5-2005

Overcoming Statistical Error and Bias in Quantum Monte Carlo: Application to Metal-Doped Helium Clusters

Gary Lee Warren Jr.
University of Tennessee, Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Gary Lee Warren Jr. entitled "Overcoming Statistical Error and Bias in Quantum Monte Carlo: Application to Metal-Doped Helium Clusters." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.

Robert J. Hinde, Major Professor

We have read this dissertation and recommend its acceptance:

Robert Harrison, Ted Barnes, Ben Xue

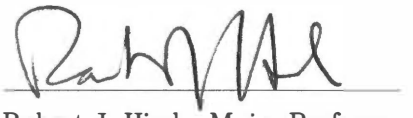
Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

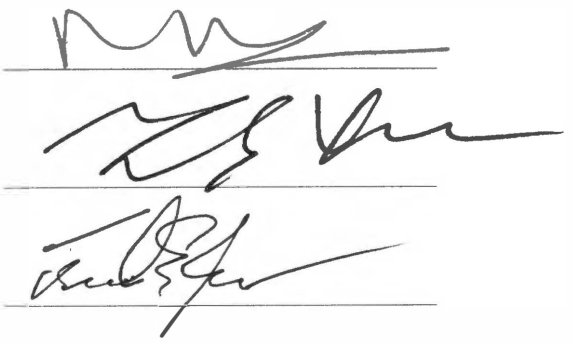
To the Graduate Council:

I am submitting herewith a dissertation written by Gary Lee Warren, Jr. entitled "Overcoming Statistical Error and Bias in Quantum Monte Carlo: Application to Metal-Doped Helium Clusters." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemistry.



Robert J. Hinde, Major Professor

We have read this dissertation
and recommend its acceptance:



Accepted for the Council:

Vice Chancellor and
Dean of Graduate Studies

OVERCOMING STATISTICAL ERROR AND BIAS IN QUANTUM MONTE CARLO: APPLICATION TO METAL-DOPED HELIUM CLUSTERS

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Gary Lee Warren, Jr.
May 2005

DEDICATION

This work is dedicated to my wife, Lena,
for her daily support and enouragement.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Robert Hinde for his model leadership as a teacher and scholar, his eye for detail (particuarly notation), and his general support and patience along the way. I would also like to thank the members of my committee, Dr. Robert Harrison for his helpful suggestions and encouragement, Dr. Ted Barnes for holding me to the standards of a "physicist", and to Dr. Ben Xue for his sincere and diverse scientific curiosity. My graditude to Dr. Jeffery Kovac, Dr. Robert Compton and Dr. John Larese for their consistent instruction, guidance, and influence throughout my graduate education.

Cheers to the past and present members of the Hinde group that I've known, Brian Taylor, Jasna Vellovic, Yosuke Kanai, Phil Stimac, Ben-hui Yang, Tim Lillestolen, James Mancillas, Cordu Samba, and Eugene dePrince for many memorable experiences and conversations.

With heartfelt gratitude I acknowledge my spouse, Lena, our parents, grandparents, and family (even the cats) for helpful encouragement and support throughout this process.

I would also like to acknowlege the National Science Foundation for generously supporting this work in the form of a Graduate Research Fellowship and Dr. John Larese for the use of computing resources.

ABSTRACT

Quantum Monte Carlo (QMC) methods are a class of powerful computer simulation techniques for solving the many-body Schrödinger equation. These techniques deliver essentially exact results and boast favorable computational scaling with system size. Calculations provide a full quantum mechanical treatment and may be carried to arbitrary precision. These characteristics make QMC a promising choice for the investigation of doped helium clusters, where quantum effects are substantial.

Stochastic in nature, QMC methods are susceptible to statistical bias and error, which must be carefully controlled. Moreover, the relationship between the finite sampling error and the statistical uncertainty in observables has never been systematically investigated. Estimates of arbitrary observables are often substandard and can be plagued by statistical uncertainties an order of magnitude or greater than those for corresponding estimates of the energy. In this work, we present an analysis of how finite populations, importance sampling, and dimensionality affect the statistical uncertainties in QMC estimates of arbitrary observables. We find that the uncertainty depends exponentially on the dimensionality of the system, independent of the observable or nature of the system. This provides insight into the minimal population sizes and importance sampling requirements necessary to obtain useful QMC estimates of properties in high-dimensional systems.

With this understanding, we develop new, more robust energy optimization procedures for cluster wavefunctions. We also implement a high quality eight parameter *ansatz* for the investigation of both pure and doped helium cluster systems. Compared to exact DMC results, the optimized wavefunctions recover over 90% of the total energy for clusters of size $n \leq 20$.

Finally, we apply this knowledge directly to the study of the solvation behavior of neutral calcium and magnesium impurities in helium nanodroplets. Diffusion Monte Carlo calculations using specially optimized Mg-He and Ca-He wavefunctions accurately determine the energetics and equilibrium structures of these highly quantum systems. We observe strong deformations of the helium density for both dopants, with calcium preferentially further from the interior of the cluster than magnesium. Finite size effects in small clusters appear to prevent the interior solvation of magnesium in clusters with 20 or fewer helium atoms. This contrasts with experimental observations in larger clusters indicating the full interior solvation of magnesium atoms.

Contents

List of Tables

List of Figures

Chapter 1

Overview

1.1 Doped cryogenic helium droplets

Cryogenic helium droplets have long been used to isolate and cool molecular species for spectroscopic investigation. Helium clusters create a nearly ideal solvation environment for atoms and molecules since the very weak interactions usually do not strongly perturb the desired spectra. The cryogenic temperatures (typically less than 0.4 K) of these droplets result in large ground state populations which greatly sharpen and simplify molecular spectra. Consequently, there is a growing interest in the experimental and theoretical characterization of cryogenic helium droplets to better understand and predict the influence of the helium environment on the spectra of embedded atoms and molecules.

Helium clusters exhibit a number of unique properties not readily described by classical arguments. For example, liquid helium has no triple-point and coexists with its vapor down to absolute zero.[1] Theoretical[2] and experimental[3,4] observations indicate that this liquid-like behavior persists even for small clusters of just a few atoms. In ^{4}He, this picture is complicated by an extremely large zero point energy and the onset of superfluidity, a frictionless liquid phase, which is readily influenced by cluster size and dopant characteristics. Spectra of embedded molecules such as HF[5], N_2O[6], and SF_6[7] have demonstrated significant finite size effects as the number of He atoms in these systems is reduced. These finite size effects in ^{4}He can radically alter the rotational constant of an embedded molecule, such as N_2O, by inducing stronger or weaker coupling with the surrounding helium atoms. In some clusters, the coupling can decrease nearly to zero, indicating the presence of a superfluid phase wherein embedded molecules rotate freely and unhindered within the cluster. The subtle interplay of finite size effects and superfluidity within these highly quantum clusters is not well understood and continues to be a driving force in current work.

In addition to using helium droplets as a spectroscopic tool for isolating and studying molecules, recent experimental investigations are exploiting these atomic and molecular impurities to interrogate the unique quantum properties of helium droplets. Neutral metal atoms, such as the alkali metals or alkaline earth metals, do not interact strongly with helium and are therefore

ideal spectroscopic probes for investigating helium nanodroplets. Recent spectroscopic studies of alkali metal dopants attached to large helium droplets[8] indicate that these impurities prefer surface sites and likely reside in small dimples on the droplet surface.[9] In surface attached configurations, these probes do not greatly perturb the structure of the cluster and are highly amenable to spectroscopic detection by virtue of their exterior location. Laser induced fluorescence (LIF) spectra of alkali metals attached to helium droplets exhibit largely unshifted $^2P_{1/2}$,$^2P_{3/2} \leftarrow {}^2S_{1/2}$ transitions with slight broadening and long tails toward higher frequencies arising from transitions into the continuum. The lack of significant broadening and negligible shifting agrees well with the proposed surface location of these dopants.

If we assume these spectral trends to be indicative of surface-attached atoms, the case for alkaline earth metals is far more puzzling. Recent LIF studies of calcium doped helium droplets with about 5000 atoms by Stienkemeier *et al.*[10] demonstrate a broad, $\sim$70 cm^{-1} blue-shifted absorption characteristic of solvated "bubble states" commonly observed in bulk helium.[11] However, the blue-shifting amounts to only a curious 40% of the estimated bulk value for calcium and even less for strontium. This may indicate that alkaline earth metals are not solvated deep within the interior of the cluster. Likewise, the apparent disagreement with the spectra of surface attached alkali atoms also seems to rule out the possibility of these atoms residing at a surface site. It is hypothesized[10] then that calcium atoms reside in deep pockets near the surface of the droplet. An equally valid possibility is that calcium atoms immersed in the highly quantum solvent make occasional excursions into the interior of the droplet, thus experiencing a variety of both interior and surface environments.

For magnesium, LIF studies of doped helium nanodroplets with around 1000 atoms conducted by Reho *et al.*[12] have resulted in the opposite conclusion, finding neutral magnesium atoms to be unambiguously solvated within the interior of the droplets. Also blue-shifted by more than 300 cm^{-1}, the magnesium $3\,{}^1P_1^o \leftarrow 3\,{}^1S_0$ transition in the droplet agrees well with the transition observed in the bulk,[13] lending credence to the interpretation that magnesium atoms are fully solvated. However, as noted by Reho *et al*, our recent knowledge of the magnesium-helium potential energy surface (PES) has not been precise enough to obtain reliable qualitative theoretical predictions about the solvation behavior of magnesium dopants within helium nanodroplets. More recently, the development of the highly accurate Mg-He and Ca-He potentials of Hinde[14] and Partridge *et al.*[15] have stimulated renewed interest in the theoretical characterization of helium clusters doped with alkaline earth metals. Concurrent work by Mella and coworkers[16] demonstrates the unique sensitivity of magnesium solvation to differences in currently available interatomic potentials. However, before discussing these results in more detail, it is worthwhile to first consider a more unified perspective on the solvation of impurities in liquid droplets.

1.2 Modeling dopant solvation in helium droplets

For a general assessment of dopant solvation within helium droplets, we turn to the λ solvation model of Ancilotto, Lerner and Cole.[17] The model is based on an analysis of the free energy

balance between attractive dopant-solvent interactions and the costly formation of a cavity within the solvent large enough to contain the impurity. This balance is characterized by a dimensionless parameter λ that is based on simple properties of the environment and interaction potential. More precisely, λ is defined as

$$\lambda \equiv 2^{-1/6} \rho \, \epsilon \, r_e / \sigma \tag{1.1}$$

where ρ is the number density of the ^{4}He solvent (0.022 Å^{-3}), ϵ and r_e are the well depth and equilibrium separation of the interparticle interaction, and σ is the surface tension of liquid ^{4}He (0.179 cm^{-1} Å^{-2}). Qualitatively, if $\lambda > 1.9$, the impurity becomes solvated within the interior of the droplet. At λ values less than 1.9, the impurity occupies a dimpled surface site. Typical values for λ vary considerably from $\lambda \simeq 0.7$ for alkali metals to $\lambda \simeq 19$ for SF_6. Alkaline earth metals are unique in the fact that their predicted λ values lie near the threshold value of 1.9. This suggests that the solvation behavior of alkaline earth atoms is very sensitive to small perturbations in the interaction potential and surrounding environment. Thus, the accurate characterization of the solvation behavior in these systems can be used to sensitively probe of the validity of the model and to aid in the development of more accurate solvation models for these systems.

For the new Mg-He and Ca-He potentials of Hinde,[14] the computed λ parameters are 2.7 and 2.1, respectively, which would favor the interior solvation of both dopants. The prediction for magnesium appears to be in agreement with current experimental findings, but the case for the interior solvation of calcium is still under investigation. Recent findings by Mella *et al.*[16] demonstrate that in small helium clusters with fewer than 30 helium atoms, finite size effects can actually shift the free energy equilibrium to the point where magnesium is expelled from the interior of the cluster, depending on the strength of the dopant-helium interaction. Consequently, with the presence of finite size effects, we are not always guaranteed that solvation will occur even with λ values significantly greater than 1.9. However, this particular effect appears extremely sensitive to the shape of the interatomic pair potential, making an accurate prediction and assessment of competing energetic factors difficult.

Considering the inherent simplicity of the λ model, one can easily envision a variety of improvements and modifications. For example, we could incorporate the zero-point energy of the dopant-solvent interaction or we could take into account more complex deformations of the helium density by using a position dependent density function, $\rho(\mathbf{r})$ in place of the isotropic density. Of course, this also neglects the importance of cooperative many-body effects which may also play a role in determining the solvation process.[6] To answer these and related questions, we are forced to employ a more thorough quantum mechanical treatment of these systems. Considering the large number of degrees of freedom present in these systems, accurate Quantum Monte Carlo (QMC) methods are well suited for these investigations.

1.3 Quantum Monte Carlo investigations of helium clusters

The application of QMC methods to the study of pure and doped helium clusters is not new, with these methods being well established techniques for the investigation of helium systems. Over the last two decades, there have been a number of theoretical investigations probing the structure and energetics of these systems. An early investigation of Pandharipande *et al.*[18] employing Green's Function Monte Carlo[19] (GFMC) determined the energy and radial density profiles of cryogenic helium clusters of various sizes. A later study by Melzer and Zabolitzky[20] confirmed the existence of no magic numbers in ^4He_n clusters and showed that the evaporation energy, $E(n-1) - E(n)$ was a smooth, monotonic function of the cluster size. This observed lack of rigidity and structure served to highlight the inherently diffuse and fluid nature of these quantum clusters.

In the last decade, a flurry of new investigations employing more sophisticated two and three body trial wavefunctions began to look more closely at structural motifs and superfluidity in these highly quantum clusters for both ground and excited states. Rick *et al.*[21] compared and contrasted the structural properties of helium clusters with those of other noble gasses. Chin and Krotscheck[22] examined structure, collective excitations, and dynamic structure factors of ^{4}He clusters. A series of papers by Whaley and coworkers[2,23–26] set new benchmarks of accuracy for QMC calculations on these systems and investigated a broad range of phenomena associated with quantum clusters of ^{4}He and H_2, including excited rotational states.

With the notable exception of Lewerenz's work on pure helium clusters,[27] most recent studies have largely focused on the properties of atomic and molecular impurities either attached to or embedded in these quantum clusters. QMC techniques have been able to successfully predict the vibrational frequency shifts of small molecules such as HF[5] and SF_6[26] embedded within helium droplets. The effects of molecular rotation of small molecules on the density of helium clusters has also been investigated[28] as well as the role of finite size effects and superfluidity on the rotational constants[6] of embedded molecules. More sophisticated QMC methods permit the examination of excited rotational states.[29]

Thus, we see that QMC methods have taken a leading role in examining the energetics, structural, and spectroscopic properties of atomic and molecular dopants embedded in quantum clusters of helium. Consequently, we will take advantage of these techniques in the present work.

1.4 Motivation

Despite the recent success of QMC techniques in helium systems, progress has been slow due to a number of factors. One problem is the current lack of highly accurate potential energy surfaces (PES) in the literature for describing dopant-helium interactions. While arbitrarily precise, QMC results for these doped clusters are no more accurate than the underlying PES. Consequently, the sustained development of reliable and accurate potential surfaces is imperative. In the present work with magnesium and calcium doped clusters, the CCSDT potentials of Hinde[14] that we employ are the most accurate surfaces currently available for these systems. Yet, despite the outstanding quality

of these potentials, the extreme sensitivity of alkaline earth metal solvation may still preclude an accurate quantitative prediction of the solvated structures.

A second issue that has limited progress in the study of helium clusters has been the construction and optimization of highly accurate and compact trial wavefunctions. Without good trial wavefunctions, convergence is slowed and the statistical error becomes problematic. For arbitrary observables, the statistical error can become unmanageably large, particularly if the boundary conditions (expecially at small separations) are not properly enforced by the form of the trial wavefunction. Thus, the use of well optimized wavefunctions is essential to obtaining accurate and reliable results. However, rigorous optimization is often difficult and time consuming, especially when the trial wavefunction contains many non-linear parameters. A number of recent advances in this area may help to relieve some of this difficulty. For the pure and doped helium systems examined in this work, we obtain excellent trial wavefunctions through the use of a well chosen *ansatz* and novel optimization procedures.

As a relatively new approach, QMC methods are still under development, both from an algorithmic as well as theoretical perspective. This has a number of important implications, the most significant one being that there exists no standardized and broadly distributed software base for these calculations. Consequently, researchers must develop and validate custom "in house" code for new investigations. Substantive resource and time investments have largely conspired to keep QMC methods specialized and underutilized tools in the sciences. Additionally, the practical and efficient computation of arbitrary properties (other than the energy) is an ongoing concern in the QMC community. Previously developed schemes are not always reliable or robust and little work has been done to characterize their accuracy and reduce the corresponding statistical error.

A major focus of this work concerns the theoretical development of QMC methods. This includes both a formal characterization of statistical error and bias in estimates of properties as well as the practical application of these ideas. In this work, we consider the statistical error as a function of sampling size, dimensionality, and importance sampling; we develop an intuitive picture through examples from the harmonic oscillator; and we apply this understanding to the optimization of trial wavefunctions and the practical estimation of properties of doped helium systems.

Physically, we are interested in characterizing the solvation behavior of alkaline earth metals attached to highly quantum helium clusters. Very accurate calculations are required to qualitatively if not quantitatively investigate the very sensitive nature of magnesium and calcium solvation in helium clusters. Consequently, we will develop high quality trial wavefunctions for these calculations, we will compute helium evaporation energies and dopant attachment energies of small He_n ($2 \leq n \leq 10$, $n = 20$) clusters and we will compare and contrast structural features of Mg- and Ca-doped clusters obtained from two-atom and three-atom distribution functions.

1.5 Outline of thesis

Chapter two provides an introduction to Quantum Monte Carlo methods, delving into the theoretical formulation of Variational Monte Carlo and Diffusion Monte Carlo methods, both of

which are used extensively throughout this work. Various aspects of these methods are highlighted and discussed within the context of statistical bias suppression and error control.

Chapter three examines the statistical error in QMC calculations of arbitrary observables and relates this error to the corresponding operator distributions. The impact of factors such as population size, dimensionality and importance sampling on the practical computation of observables are elucidated.

Chapter four directly applies the theoretical developments of chapter three to the k-dimensional harmonic oscillator—a preliminary approximation of high-dimensional cluster systems. Trends relating to population size, dimensionality, and importance sampling are examined for specific operators.

Chapter five builds on our understanding of the statistical error expression in doped helium clusters by considering new procedures for the optimization of pure and doped helium cluster wavefunctions. Optimized wavefunctions are developed for He-He, Mg-He, and Ca-He interactions in clusters of varying size. The effectiveness of a new trial wavefunction *ansatz* and the quality of the resulting optimized trial functions is assessed.

Chapter six applies the optimized wavefunctions developed in chapter five to Diffusion Monte Carlo investigations of magnesium and calcium doped helium clusters. Total energetics, helium evaporation energies, and dopant binding energies are computed. Structural information obtained from two- and three-particle correlation functions helps to identify structural motifs and trends present in the solvation of magnesium and calcium atoms.

Chapter seven provides a summary of new theoretical ideas presented in this dissertation and important results from magnesium and calcium doped systems. Ongoing work and future applications are also discussed.

Chapter 2

Overview of Quantum Monte Carlo

2.1 Introduction

Having outlined our primary interest in the theoretical characterization of atomic dopants in cryogenic helium clusters, we now provide a description of the theoretical methods that we will employ in this investigation. Quantum Monte Carlo methods will be our tools of choice because of their ability to accurately treat large many-body quantum systems. In this work, we will employ two distinct types of QMC methods, *Variational Monte Carlo* (VMC) and *Diffusion Monte Carlo* (DMC). As complementary approaches, these two techniques posses different strengths and weaknesses. VMC is simpler and faster but less accurate than DMC which is more complex and essentially exact.

In this chapter, we review the basic theory of quantum Monte Carlo methods, starting with simple Monte Carlo integration. From there, we discuss Markov chains and their role in the construction and sampling of discrete distributions. Next, we consider various theoretical and practical aspects of VMC and DMC algorithms. Finally, we conclude by discussing sources of error and bias present in practical implementations of both algorithms along with methods for error suppression and control.

2.2 Monte Carlo integration

First described by Metropolis and Ulam[30] in 1949, the "Monte Carlo" method is a computational tool for the numerical integration of arbitrary functions. It is a *stochastic* approach that uses random numbers to distribute sampling points over an integration region. Integrated properties are computed as averages over these sampling points and are subject to a statistical uncertainty that shrinks as the number of sampling points is increased. Though only recently applied to the solution of problems in molecular physics,[31] the method itself has been in use for far longer, having been employed by Buffon[32] in 1777 to compute the value of π in his famous needle experiment. However, it is only with the advent of modern computing and rapid random number generation that the Monte

Carlo method has enjoyed widespread use.

Compared with most other methods of numerical integration such as Gaussian quadrature, Monte Carlo integration is strikingly inefficient. Moreover, the error in standard methods of integration is bounded and dictated by the grid spacing. In contrast, Monte Carlo procedures can only yield an error *estimate* within some level of statistical confidence. This means that unless the sampling is infinite, we cannot even guarantee that the true result lies within our confidence limits. This would strongly deter anyone from using Monte Carlo techniques were it not for the fact that all other integration methods are computationally intractable in high dimensional space. Weighing an approximate solution against no solution at all, the Monte Carlo method then becomes a reasonable compromise for computing multi-dimensional integrals.

Grid based numerical integration methods are *deterministic*. The only error in deterministic numerical integration is a discretization error that depends on the spacing and number of grid points. Finer grids reduce this error by permitting a more faithful representation of continuous integrands. However, the error reduction achieved by refining the grid comes with increased computational cost, becoming exceedingly expensive in higher dimensions. If, for example, we consider a k dimensional function and an integration grid of 100 points per dimension, the number of function evaluations scales as 100^k. This exponential scaling of computational cost becomes prohibitive for even a relatively small number of dimensions. More efficient schemes (like Hermite-Gauss quadrature) that incorporate weighting can deliver comparable errors with fewer grid points, but the exponential "curse of dimensionality" [33] is inescapable. In high dimensions, these methods have the added drawback that the grid can only be expanded stepwise. Extending a k dimensional grid along just one dimension increases the grid by N^{k-1} points, which can be quite demanding.

In contrast, Monte Carlo integration is a practical alternative to computing high dimensional integrals. In Monte Carlo integration, the integration grid is replaced with a small sampling population. This population may be as small as a single point or include many thousands of sampling points. These points are dispersed throughout the integration space and their locations vary with time. The sampling population at a specific moment in time is referred to as a *population snapshot*. Each snapshot provides an estimate of the integral or integrated property. These estimates will fluctuate from snapshot to snapshot with a variance dependent on the sampling population size N and proportional to N^{-1}. The error of the mean of these estimates is proportional to the square root of the variance, or $N^{-1/2}$, given the assumption that the samples are *independent* or *uncorrelated*.

Monte Carlo integration has the advantage of being able to explore integration space "more densely" than grid based methods. Since sampling points are unrestricted in location, they may explore the space between grid points that is otherwise inaccessible to deterministic methods. This can be important if the grid spacing is too coarse or if the integrand possesses high local curvature. The sampling population in Monte Carlo can also be expanded to include an arbitrary number of sampling points. Even if we take our sampling population to consist of just one sampling point, we still may obtain the mean value for the integral by averaging over many snapshots. To obtain an estimate of the integral, this single sampling point must have the opportunity to explore all of the integration space, or at least all of the *important* integration space where the integrand contributes.

Employing a larger population size reduces the statistical error since larger populations "better" sample the integrand at each snapshot. For example, using a sampling population ten times larger would reduce the variance from snapshot to snapshot by a factor of ten. This flexibility leads to a tunable balance between computational resources and overall execution time. Moreover, the statistical independence of sampling points strongly favors the design of parallel algorithms optimal for both Beowulf-style computing clusters, shared memory systems, and potentially the next generation of multi-core cell processors.

The standard Monte Carlo integration algorithm is concise and straightforward to implement. For example, to compute the integral or expected value

$$\int g(\mathbf{x})f(\mathbf{x})d\mathbf{x}, \tag{2.1}$$

of the weighting function $g(\mathbf{x})$ over $f(\mathbf{x})$ where $\mathbf{x}$ is a k dimensional state vector, we would use the following algorithm:

1. Generate a new population of sampling points of size N distributed according to $f(\mathbf{x})$.
2. For each sampling point $\mathbf{x}_i$, evaluate and accumulate values of $g(\mathbf{x}_i)$.
3. Compute the average value, $\bar{g}$, of $g(\mathbf{x})$ over all sampling points in the population.
4. Repeat steps 1-3 many times, accumulating values of $\bar{g}$ and $\bar{g}^2$ for each snapshot.
5. Compute the ensemble average and variance of $\bar{g}$ over all snapshots.

The computation of these averages is trivial if we already have sampling points distributed according to $f(\mathbf{x})$. However, if $f(\mathbf{x})$ is complicated or unknown (as is often the case) we require a method for transforming an arbitrarily distributed population into a new population distributed according to $f(\mathbf{x})$. We accomplish this through the use of Markov chains.

2.3 Markov Chains

A Markov chain is a stochastic (random) process wherein state to state evolution is influenced only by the current state.[34] In this manner, a Markov process is memoryless and evolves independently of its history. Mathematically, the Markov property is expressed as

$$P(\mathbf{X}_{n+1} = \mathbf{x}_{n+1}|\mathbf{X}_0 = \mathbf{x}_0, \ldots, \mathbf{X}_n = \mathbf{x}_n) = P(\mathbf{X}_{n+1} = \mathbf{x}_{n+1}|\mathbf{X}_n = \mathbf{x}_n) \tag{2.2}$$

where the conditional probability that the random variable $\mathbf{X}_{n+1}$ takes on the value $\mathbf{x}_{n+1}$ depends only on the fact that $\mathbf{X}_n = \mathbf{x}_n$ and not on the values of any other random variables in the chain. The quantity $P(\mathbf{X}_{n+1} = \mathbf{x}_{n+1}|\mathbf{X}_n = \mathbf{x}_n)$ is called a transition probability and is often denoted as $P(\mathbf{X}_n \rightarrow \mathbf{X}_{n+1})$. A Markov process is commonly described in terms of a random walk, in which the values $\mathbf{x}_{n+1}$ are actual positions and the set of all $\mathbf{x}_i$ forms the path. In such a walk, moves are only

influenced by information about the present location. This is in contrast to a self-avoiding random walk (non-Markovian) which depends on the entire path.

For ergodic Markov chains, there exists a steady state or stationary distribution, $\pi(\mathbf{x})$, to which the process converges after infinitely many steps. The ergodic property requires that all states are accessible with non-zero transition probabilities, that the transitions are not periodic, and that states are recurrent. If these conditions are met, the Markov chain evolves into an equilibrium or stationary distribution that is the time invariant "solution" of the Markov chain. After infinitely many transitions we have

$$\pi(\mathbf{x}_{n+1}) = \int P(\mathbf{x}_n \to \mathbf{x}_{n+1})\, \pi(\mathbf{x}_n)\, d\mathbf{x}_n \tag{2.3}$$

and additional transitions simply replicate the distribution $\pi(\mathbf{x})$. Thus, integration over all possible transitions to other states results in the same distribution of states. For conservation of probability to hold, the transition probabilities must be bounded according to

$$0 \leq P(\mathbf{x}_n \to \mathbf{x}_{n+1}) \leq 1 \tag{2.4}$$

and normalized so that

$$1 = \int P(\mathbf{x}_n \to \mathbf{x}_{n+1})\, d\mathbf{x}_n. \tag{2.5}$$

Within the context of Monte Carlo integration, Markov chains can be used to construct a sampling population distributed according to the desired integrand, $f(\mathbf{x})$. This entails engineering a Markov chain that converges to the stationary distribution $\pi(\mathbf{x}) = f(\mathbf{x})$. One way of ensuring this is to invoke detailed balance on the transition probabilities. For the distribution to be stationary, we know that overall, transitions to some point $\mathbf{x}$ must balance transitions away from $\mathbf{x}$ so that there is no net change to $\pi(\mathbf{x})$:

$$\int P(\mathbf{x} \to \mathbf{x}')\, \pi(\mathbf{x})\, d\mathbf{x} = \int P(\mathbf{x}' \to \mathbf{x})\, \pi(\mathbf{x}')\, d\mathbf{x}'. \tag{2.6}$$

The strict detailed balance condition

$$P(\mathbf{x} \to \mathbf{x}')\, \pi(\mathbf{x}) = P(\mathbf{x}' \to \mathbf{x})\, \pi(\mathbf{x}') \tag{2.7}$$

forces this to be true for individual transitions while at the same time satisfying equation 2.6. This is a sufficient condition to guarantee convergence in an ergodic Markov chain. If we now rewrite the transition probability as

$$P(\mathbf{x} \to \mathbf{x}') = A(\mathbf{x} \to \mathbf{x}')\, T(\mathbf{x} \to \mathbf{x}') \tag{2.8}$$

in terms of an *acceptance* probability $0 \leq A(\mathbf{x} \rightarrow \mathbf{x}') \leq 1$ and a transition function $T(\mathbf{x} \rightarrow \mathbf{x}')$, we can recast the strict detailed balance equation as

$$\frac{A(\mathbf{x} \rightarrow \mathbf{x}')}{A(\mathbf{x}' \rightarrow \mathbf{x})} = \frac{T(\mathbf{x}' \rightarrow \mathbf{x})\,\pi(\mathbf{x}')}{T(\mathbf{x} \rightarrow \mathbf{x}')\,\pi(\mathbf{x})}. \tag{2.9}$$

Application of the strict detailed balance condition fixes the ratio of acceptance probabilities but still leaves the probabilities themselves undetermined. Consequently, we have some freedom of choice in the selection of acceptance probabilities. The Metropolis method makes the choice $A(\mathbf{x}' \rightarrow \mathbf{x}) = 1$. Since $A(\mathbf{x} \rightarrow \mathbf{x}')$ is now relative to $A(\mathbf{x}' \rightarrow \mathbf{x})$, the forward transition is accepted if the right hand side is greater than unity. If less than unity, the transition is conditionally accepted with a probability given by the right hand side. This is a very convenient choice for the acceptance probability since it makes it unnecessary to explicitly determine the normalization of $\pi(\mathbf{x})$.

A more succinct expression of the Metropolis forward acceptance probability is

$$A(\mathbf{x} \rightarrow \mathbf{x}') = \min\left\{1, \frac{T(\mathbf{x}' \rightarrow \mathbf{x})\,\pi(\mathbf{x}')}{T(\mathbf{x} \rightarrow \mathbf{x}')\,\pi(\mathbf{x})}\right\}. \tag{2.10}$$

In many cases, the transition function $T(\mathbf{x} \rightarrow \mathbf{x}')$ is symmetric with respect to $\mathbf{x}$ and $\mathbf{x}'$ and will cancel completely from equation 2.10, leaving only the ratio of $\pi(\mathbf{x}')$ to $\pi(\mathbf{x})$. With this choice of the acceptance probability, the Metropolis procedure will generate configurations distributed according to $\pi(\mathbf{x})$. Depending on the specific quantum Monte Carlo method, $\pi(\mathbf{x})$ is proportional to either a nodeless wavefunction $\Psi(\mathbf{x})$, or the corresponding probability density $|\Psi(\mathbf{x})|^2$.

2.4 Variational quantum Monte Carlo

2.4.1 Introduction

Variational Monte Carlo is a method for numerically computing quantum mechanical expectation values for arbitrary trial wavefunctions in many-dimensional systems. As the name suggests, the term "variational" comes from the variational principle which is used to obtain upper bounds on quantities such as the ground state energy. In VMC, trial wavefunctions incorporate a set of adjustable parameters that are varied in an attempt to minimize these upper bounds. The optimal parameters yield a trial wavefunction $\Psi_T(\mathbf{x})$ of a given functional form that "best" approximates the exact wavefunction, $\Psi_0(\mathbf{x})$.

In VMC, we use a Markov chain and the Metropolis method to construct the probability density $\rho(\mathbf{x}) = |\Psi_T(\mathbf{x})|^2$. Standard Monte Carlo integration is then used to extract meaningful expectation values of the wavefunction. For coordinate dependent quantities $A(\mathbf{x})$, expectation values are given by the ensemble average

$$\langle A \rangle = \frac{\int \rho(\mathbf{x}) A(\mathbf{x}) d\mathbf{x}}{\int \rho(\mathbf{x}) d\mathbf{x}} \tag{2.11}$$

over the system.

This method is most efficient for quantum many-body systems in which the number of dimensions is large and analytical integration is impossible. A key advantage of VMC over analytical integration concerns the complexity of the trial wavefunction. In VMC, since we are performing numerical integration, complicated functional forms for $\Psi_T(\mathbf{x})$ may be employed to generate a compact representation of the wavefunction.[35] This offers certain advantages over traditional *ab initio* approaches in which wavefunctions are expanded in (often rather large) sets of basis functions. Such compact representations often better describe the essential physics, require fewer parameters and contribute to better scaling of the VMC algorithm. Monte Carlo methods are generally considered to scale computationally between n^2 and n^3 with the number of particles (n) in the system[35] which is quite good in comparison with the scaling of other accurate *ab initio* techniques.

Perhaps one of the most disturbing features of VMC is the statistical uncertainty which depends on the quality of $\Psi_T(\mathbf{x})$. High accuracy calculations should be able to detect small differences in the quality of trial wavefunctions. However, these small differences are often dwarfed by the statistical error in a calculation. In general, the statistical error scales as $N^{-1/2}$ where N is the number of statistical samples. Through $\Psi_T(\mathbf{x})$, the statistical error can also depend on the dimensionality of the system. Empirical evidence suggests that the statistical error rises exponentially with the dimensionality of the system, posing serious computational challenges for optimizing multidimensional wavefunctions to high accuracy. This dependence on dimension will be addressed in more detail in chapter 3.

2.4.2 Expectation values and statistical correlation

Expectation values in VMC are obtained as simple ensemble averages over a suitable probability distribution. This probability distribution must be proportional to $|\Psi_T(\mathbf{x})|^2$ and is commonly constructed and sampled with the Metropolis, Fokker-Planck, or generalized Metropolis algorithms described below. Instantaneous estimates of coordinate operators are given as averages over the sampling population

$$A_N = \frac{\int |\Psi_T(\mathbf{x})|^2 A(\mathbf{x}) d\mathbf{x}}{\int |\Psi_T(\mathbf{x})|^2 d\mathbf{x}} \cong \frac{1}{N} \sum_{i=1}^{N} A(\mathbf{x}_i). \tag{2.12}$$

Likewise, derivative operators may be sampled as local operators $\Psi_T(\mathbf{x})^{-1} \hat{A} \Psi_T(\mathbf{x})$ according to

$$A_N = \frac{1}{\int |\Psi_T(\mathbf{x})|^2 d\mathbf{x}} \int |\Psi_T(\mathbf{x})|^2 \left(\frac{\hat{A}\Psi_T(\mathbf{x})}{\Psi_T(\mathbf{x})} \right) d\mathbf{x} \cong \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{A}\Psi_T(\mathbf{x}_i)}{\Psi_T(\mathbf{x}_i)}. \tag{2.13}$$

If these instantaneous estimates are statistically independent, we could treat each A_N as a snapshot and simply average a large number of them to obtain the simulation average. However, successive or serial samplings of $|\Psi_T(\mathbf{x})|^2$ are almost always correlated and do not produce fully independent estimates. Correlation is a measure of the "relatedness" between successive samplings. Perfect correlation would be obtained if we repeatedly reused the same sampling points to generate

instantaneous estimates. In such a case, the lack of randomness would result (falsely) in no statistical error and an incorrect simulation average. Correlation can be considered a type of memory in which the *correlation time* is representative of how long it takes to "forget" information about previously used sampling points. The correlation time τ_c in sequential estimates of an operator $\hat{A}$ can formally be defined as

$$\tau_c = 1 + 2\sum_{\ell=1}^{\ell_0} C_A(\ell) \tag{2.14}$$

where $C_A(\ell)$ is the autocorrelation function

$$C_A(\ell) = \frac{1}{(m-\ell)} \sum_{j=1}^{m-\ell} \frac{(A_{N,j} - \langle A_N \rangle)(A_{N,j+\ell} - \langle A_N \rangle)}{\langle A_N^2 \rangle - \langle A_N \rangle^2}. \tag{2.15}$$

In the above formulas, ℓ is the *lag time* or the number of instantaneous estimates separating the jth and $(j+\ell)$th estimates and ℓ_0 is the lag time for which the autocorrelation function becomes zero.

To eliminate unwanted correlations from our data, we must invoke a blocking procedure, whereby $N_B >= \tau_c$ instantaneous estimates are grouped together into a block and averaged. This produces a single, independent block/snapshot estimate $A_{N,B}$ where

$$A_{N,B} \cong \frac{1}{NN_B} \sum_{j=1}^{N_B} \sum_{i=1}^{N} A_{i,j}. \tag{2.16}$$

We can now define the variational or VMC expectation value $\langle A \rangle_v$ over the simulation as an average over M blocks,

$$\langle A \rangle_v \cong \frac{1}{M} \sum_{i=1}^{M} A_{N,B_i}. \tag{2.17}$$

Judicious application of this block averaging procedure will effectively remove the correlation bias in practical calculations. However, because we have reduced the number of independent observations in our average, the statistical sampling variance

$$s_A^2 = \frac{1}{M-1} \left(\langle A_{N,B}^2 \rangle - \langle A_{N,B} \rangle^2 \right) \tag{2.18}$$

for estimates of the operator $\hat{A}$ over the simulation will now be larger than if we had neglected the blocking procedure. This means that more instantaneous estimates must be generated to maintain the same statistical error bars. We can write this increase in variance in terms of a *statistical inefficiency* S as

$$S = \lim_{N_B \to \infty} N_B \frac{s_A^2}{\tilde{s}_A^2}, \tag{2.19}$$

where $\tilde{s}_A^2$ is the variance of the unblocked data. As the block size N_B increases, the statistical inefficiency will plateau.

2.4.3 Sampling algorithms

A. Standard Metropolis algorithm

In the standard Metropolis approach, we seek to construct a Markov chain with the probability distribution $|\Psi_T(\mathbf{x})|^2$ as the asymptotic stationary distribution. This is accomplished through the use of an acceptance/rejection step, in which certain sampling points are discarded with some probability. To assemble the Markov chain, we begin with a population of sampling points with coordinates $\mathbf{x}_i$. We displace these points within configuration space according to a set of uniformly distributed random vectors, $\vec{\xi}_i \in [-1, 1]$. These random vectors are scaled by some tunable constant η, known as the box size, and are added to positions of the current sampling points to yield the new sampling points, $\mathbf{x}_i'$. The acceptance probability

$$A(\mathbf{x}_i \to \mathbf{x}_i') = \min\left\{1, \frac{|\Psi_T(\mathbf{x}_i')|^2}{|\Psi_T(\mathbf{x}_i)|^2}\right\} \tag{2.20}$$

is evaluated for each displacement. If $A(\mathbf{x}_i \to \mathbf{x}_i') \geq 1$ then the move is retained; otherwise, we compare $A(\mathbf{x}_i \to \mathbf{x}_i')$ to a new uniform random number ζ between 0 and 1. If $A(\mathbf{x}_i \to \mathbf{x}_i') \geq \zeta$, we keep the move. Only in the case $A(\mathbf{x}_i \to \mathbf{x}_i') < \zeta$ do we reject the move to $\mathbf{x}_i'$ and retain the sampling point $\mathbf{x}_i$. The final set of sampling points, including both the new ones that were accepted and the old ones resulting from rejections, comprise the new sampling population. For computational efficiency, the box size is adjusted to tune the percentage of rejected moves to around 50%. Larger box sizes result in faster equilibration at the expense of more rejections and accordingly, more correlation. Smaller box sizes yield small displacements which also result in increased correlation between sampling points.

If we repeatedly generate new sampling points, after many iterations the sampling population approaches its asymptotic limit of $|\Psi_T(\mathbf{x})|^2$. Ensemble averages of observable operators may be collected over additional iterations. The larger the equilibrated sampling population, the smaller the statistical error of computed averages.

B. Fokker-Planck algorithm

A more computationally efficient approach is the Fokker-Planck scheme wherein we simulate a differential equation constructed to have the real steady-state solution Ψ_T^2. For this we use the Fokker-Planck equation

$$\frac{\partial f(\mathbf{x}, \tau)}{\partial t} = D\,\nabla^2 f(\mathbf{x}, \tau) - D\,\nabla \cdot \left[f(\mathbf{x}, \tau)\nabla \ln \Psi_T^2(\mathbf{x})\right] \tag{2.21}$$

describing the time evolution of a distribution function $f(\mathbf{x}, \tau)$. The form of the Fokker-Planck equation resembles the Schrödinger and diffusion equations, but with an added drift term. Here, D is the diffusion constant equal to $\hbar^2/2m$ and the quantity $\nabla \ln \Psi_T^2(\mathbf{x})$ is often referred to as the *quantum force* or *quantum velocity*, $\mathbf{F}(\mathbf{x})$. The drift term "adjusts" the pure diffusion process to produce Ψ_T^2 as the stationary distribution. Instead of coupling a random walk model with a rejection

scheme as in the Metropolis algorithm, we can generate Ψ_T^2 directly with the directed random walk

$$\mathbf{x}' = \mathbf{x} + D\Delta\tau\mathbf{F}(\mathbf{x}) + \sqrt{2D\Delta\tau}\,\vec{\xi}. \tag{2.22}$$

New sampling points are generated by applying a random Gaussian displacement $\vec{\xi}$ of zero mean and unit variance and adding a drift proportional to the quantum force. The drift encourages sampling points to diffuse or walk into regions where Ψ_T^2 is large, thereby increasing the efficiency of the algorithm relative to the standard Metropolis approach. Unlike the Metropolis method, no acceptance/rejection step is required to generate the stationary distribution. The step size ($\Delta\tau$) controls the time discretization of the process.

In practice, the Fokker-Planck process is only exact in the limit of $\Delta\tau \to 0$ because we only simulate the Fokker-Planck equation to 1st order in the difference $\Delta\tau$.[36] Thus any finite value of $\Delta\tau$ will produce a stationary distribution slightly different from $\Psi_T^2(\mathbf{x})$. This problem can be suppressed or eliminated in several ways. First, we can simulate the Fokker-Planck equation to higher order using more sophisticated Runge-Kutta type algorithms. However, the higher the order, the more work involved to take a single step, which eventually defeats the efficiency of the method. Second, we can perform simulations at various values of $\Delta\tau$ and extrapolate properties to $\Delta\tau = 0$ in an effort to eliminate the bias. Again, this involves more work, but is a reliable method of eliminating the bias. A third approach is to couple the Fokker-Planck algorithm with the Metropolis scheme into a hybrid method, the generalized Metropolis algorithm.

C. Generalized Metropolis algorithm

In the generalized Metropolis algorithm, we seek to remedy the finite time step bias in the Fokker-Planck algorithm by rejecting some sampling points, such that the equilibrium distribution always converges to $\Psi_T^2(\mathbf{x})$, regardless of the value of $\Delta\tau$. A sufficient condition to guarantee this convergence is to invoke detailed balance. Incorporating the strict detailed balance condition (equation 2.7) into a Metropolis scheme, we must accept or reject Fokker-Planck moves according to

$$A(\mathbf{x} \to \mathbf{x}') = \min\left\{1, \frac{|\Psi_T(\mathbf{x}')|^2\, T(\mathbf{x}' \to \mathbf{x}; \Delta\tau)}{|\Psi_T(\mathbf{x})|^2\, T(\mathbf{x} \to \mathbf{x}'; \Delta\tau)}\right\} \tag{2.23}$$

where the forward finite-time transition function $T(\mathbf{x} \to \mathbf{x}'; \Delta\tau)$ is given by the shifted Gaussian

$$T(\mathbf{x} \to \mathbf{x}'; \Delta\tau) = (4\pi D\Delta\tau)^{k/2} e^{\left(\mathbf{x}' - \mathbf{x} - D\Delta\tau F(\mathbf{x})\right)^2 / 4D\Delta\tau}, \tag{2.24}$$

which is a short time approximation to the Green's function solution of the Fokker-Planck equation.

Typically, the time step is tuned so that the number of rejected moves is quite small, usually less than 0.1%.[27,37] Larger values of the time step increase the percentage of rejected moves. Both small and large values of the time step tend to introduce unwanted correlation effects. For small $\Delta\tau$, many more steps are required to obtain statistically independent configurations. For large $\Delta\tau$, a significant number of sampling configurations can be repeatedly rejected, resulting in stagnant or

persistent configurations that disproportionately contribute to the statistics.[37] To help suppress bias from too many rejected configurations, we employ the delayed rejection procedure described below.

D. Delayed rejection algorithm

Delayed rejection, first introduced in Markov Chain Monte Carlo (MCMC) by Mira and Tierney[38–40] and recently applied to QMC methods by Bressanini and coworkers,[41] is a modification for decreasing correlation effects due to rejected moves within the standard and generalized Metropolis algorithms. In the delayed rejection algorithm, we propose an alternate move in the event an initial move is rejected. We take advantage of information obtained from the rejected first move to make the process more efficient, but alter the second-stage acceptance probability to preserve the convergence of the Markov chain.

In the standard Metropolis approach, we implement delayed rejection by first attempting a move with displacement $\mathbf{m}$ according the standard acceptance probability

$$A_1(\mathbf{x} \overset{\mathbf{m}}{\rightarrow} \mathbf{y}) = \min\left\{1, \frac{|\Psi_T(\mathbf{y})|^2}{|\Psi_T(\mathbf{x})|^2}\right\}. \tag{2.25}$$

Only if this move is rejected do we attempt to construct the acceptance probability for an additional move to a new location, $\mathbf{z}$. One method of ensuring that the stationary distribution will remain unaffected by the second move is to consider a hypothetical move from $\mathbf{z}$ to some other configuration $\mathbf{y}'$ with the displacement $(-\mathbf{m})$. This relates $\mathbf{y}'$ to $\mathbf{y}$ through a kind of inversion symmetry about the move from $\mathbf{x}$ to $\mathbf{z}$. The acceptance probability for this move is given by

$$A_2(\mathbf{z} \overset{-\mathbf{m}}{\rightarrow} \mathbf{y}') = \min\left\{1, \frac{|\Psi_T(\mathbf{y}')|^2}{|\Psi_T(\mathbf{x})|^2}\right\}. \tag{2.26}$$

We may now compute the acceptance probability for the proposed alternate move from $\mathbf{x}$ to $\mathbf{z}$ via the modified acceptance probability

$$A(\mathbf{x} \rightarrow \mathbf{z}) = \min\left\{1, \left(\frac{1 - A_2(\mathbf{z} \overset{-\mathbf{m}}{\rightarrow} \mathbf{y}')}{1 - A_1(\mathbf{x} \overset{\mathbf{m}}{\rightarrow} \mathbf{y})}\right) \frac{|\Psi_T(\mathbf{z})|^2}{|\Psi_T(\mathbf{x})|^2}\right\}. \tag{2.27}$$

If this second move is accepted, we take $\mathbf{z}$ to be the new sampling point, otherwise the sampling point remains at $\mathbf{x}$. The delayed rejection algorithm may be extended to incorporate additional attempts if the second move is also rejected. However, additional delayed rejection moves are often subject to diminishing returns with respect to computational efficiency.

Coupling a single delayed rejection step with the generalized Metropolis algorithm is desirable, since it permits a larger time step to be employed while suppressing the correlation. The delayed rejection implementation for the combined Fokker-Planck and generalized Metropolis algorithm is complicated by the added drift term in the diffusion. As with the standard algorithm, we

first apply the drift and a Gaussian displacement to generate a new configuration $\mathbf{y}$ according to

$$\mathbf{y} = \mathbf{x} + D\Delta\tau\mathbf{F}(\mathbf{x}) + \sqrt{2D\Delta\tau}\vec{\xi}_1 \tag{2.28}$$

with acceptance probability

$$A_1(\mathbf{x} \to \mathbf{y}) = \min\left\{1, \frac{|\Psi_T(\mathbf{y})|\ T(\mathbf{y} \to \mathbf{x})}{|\Psi_T(\mathbf{x})|\ T(\mathbf{x} \to \mathbf{y})}\right\}. \tag{2.29}$$

If this move is rejected, we construct a new move from $\mathbf{x}$ to $\mathbf{z}$ and the equivalent move from $\mathbf{z}$ to $\mathbf{y}'$, mirroring the diffusion displacement $\vec{\xi}_1$, but employing the quantum force at $\mathbf{z}$ according to

$$\mathbf{z} = \mathbf{x} + D\Delta\tau\mathbf{F}(\mathbf{x}) + \sqrt{2D\Delta\tau}\vec{\xi}_2. \tag{2.30}$$

$$\mathbf{y}' = \mathbf{z} + D\Delta\tau\mathbf{F}(\mathbf{z}) + \sqrt{2D\Delta\tau}(-\vec{\xi}_1). \tag{2.31}$$

The mirrored move has acceptance probability

$$A_2(\mathbf{z} \to \mathbf{y}') = \min\left\{1, \frac{|\Psi_T(\mathbf{y}')|^2\ T(\mathbf{y}' \to \mathbf{z})}{|\Psi_T(\mathbf{z})|^2\ T(\mathbf{z} \to \mathbf{y}')}\right\}, \tag{2.32}$$

and the final acceptance probability for the proposed second move to $\mathbf{z}$ is then

$$A(\mathbf{x} \to \mathbf{z}) = \min\left\{1, \left(\frac{1 - A_2(\mathbf{z} \to \mathbf{y}')}{1 - A_1(\mathbf{x} \to \mathbf{y})}\right) \frac{|\Psi_T(\mathbf{z})|^2\ T(\mathbf{z} \to \mathbf{x})}{|\Psi_T(\mathbf{x})|^2\ T(\mathbf{x} \to \mathbf{z})}\right\}. \tag{2.33}$$

The calculation of this acceptance probability is computationally expensive, requiring four evaluations of the quantum force. However, it is executed only when the first move is rejected—a relatively rare occurrence in the generalized Metropolis algorithm where first-move rejection rates are usually less than 0.1%.

2.5 Diffusion Monte Carlo

2.5.1 Introduction

Diffusion Monte Carlo[30,42,43] which is complementary to VMC is a technique for numerically solving the many-body Schrödinger equation. Diffusion or "projector" Monte Carlo employs the imaginary time evolution operator to project the exact ground state wavefunction out of an arbitrary initial distribution. Averages and integrated properties are then computed over this ground state wavefunction. In DMC, the ground state wavefunction is not known *a priori*, but generated *in situ*; this contrasts with VMC in which the quality of a *known* trial wavefunction is assessed and optimized within a fixed parameter set.

Diffusion Monte Carlo comes in two varieties, Pure Diffusion Monte Carlo (PDMC) which samples a wavefunction and Importance Sampled Diffusion Monte Carlo (ISDMC) that samples

a probability density. PDMC is a relatively simple algorithm to implement but suffers from large statistical noise and slow convergence which have deterred its widespread application. Consequently, most modern DMC methods also implement *importance sampling*, which employs a variationally optimized trial function to accelerate convergence and suppress statistical fluctuations. This benefit of well optimized trial functions underscores the importance of variational Monte Carlo methods within the context of diffusion Monte Carlo. In this work, we will employ importance sampling with variationally optimized wavefunctions for all production level DMC simulations of atomic clusters.

2.5.2 Theoretical Discussion of the DMC Algorithm

A. Time transformation and projection

The DMC method relies on an isomorphism between the diffusion or heat equation

$$\frac{\partial u(x,t)}{\partial t} = k\nabla^2 u(x,t) - q(x)u(x,t) \tag{2.34}$$

and the time-dependent Schrodinger equation written in imaginary time

$$\frac{\partial \Psi(x,\tau)}{\partial \tau} = \frac{\hbar^2}{2m}\nabla^2 \Psi(x,\tau) - V(x)\Psi(x,\tau). \tag{2.35}$$

Here, we have used the transformation $\tau = it/\hbar$, also known as a Wick rotation, to cast the Schrödinger equation in this form. With this time transformation, stationary eigenstates which have a time-dependent phase factor

$$\phi_n(x)e^{-itE_n/\hbar} \tag{2.36}$$

become transient states

$$\phi_n(x)e^{-\tau E_n} \tag{2.37}$$

decaying exponentially in imaginary time with rates specified by their energy eigenvalues. We can expand any solution of equation 2.35 at time τ in terms of these transients as

$$\Psi(x,\tau) = \sum_{n=0}^{\infty} c_n \phi_n(x) e^{-\tau E_n} \tag{2.38}$$

where the expansion coefficients are given by the overlap integral

$$c_n = e^{\tau_0 E_n} \int_{-\infty}^{\infty} \phi_n^*(x_0)\Psi(x_0,\tau_0)\, dx_0. \tag{2.39}$$

As we advance forward in imaginary time, states with positive energy eigenvalues will decay and states with negative eigenvalues will grow. Since we are interested in properties of the ground state, it would convenient if we could stabilize the ground state with respect to other states. This is accomplished by introducing a constant energy shift in the Schrödinger equation. If we replace $V(x)$

with $V(x) - E_R$ and E_n with $E_n - E_R$, the resulting solutions

$$\phi_n(x)e^{-it(E_n-E_R)/\hbar} \tag{2.40}$$

allow us expand an arbitrary state as

$$\Psi(x,\tau) = \sum_{n=0}^{\infty} c_n\phi_n(x)e^{-\tau(E_n-E_R)}. \tag{2.41}$$

If we now choose $E_R = E_0$, the ground state component will remain finite while all other states decay in imaginary time, assuming that c_0 was non-zero initially. In this manner, DMC can project a static, fixed ground state component out of an arbitrary initial wavefunction.

B. Propagation and evolution operators

The process of carrying this projection process forward in time is referred to as *propagation.* Using equation 2.39, we can rewrite equation 2.41 in the integral form

$$\Psi(x,\tau) = \int_{-\infty}^{\infty} \mathcal{K}(x,\tau;x_0,\tau_0)\Psi(x_0,\tau_0)\,dx_0 \tag{2.42}$$

where the kernel or *propagator*, $\mathcal{K}(x,\tau;x_0,\tau_0)$, is defined as

$$\mathcal{K}(x,\tau;x_0,\tau_0) = \sum_{n=0}^{\infty} \phi_n^*(x_0)\phi_n(x)e^{-\Delta\tau(E_n-E_R)} \tag{2.43}$$

and where $\Delta\tau = \tau - \tau_0$. The propagator describes the influence that some point x_0 at time τ_0 has on the wavefunction at a later time τ. Integrated over all x_0, $\mathcal{K}(x,\tau;x_0,\tau_0)$ determines the wavefunction at time τ given the wavefunction at time τ_0. The role of the propagator in diffusion Monte Carlo is to evolve the wavefunction forward in time, effectuating the projection process. However, for practical problems, the true propagator is seldom known. Moreover, equation 2.43 defines $\mathcal{K}(x,\tau;x_0,\tau_0)$ in terms of the very eigenfunctions that we are seeking. Thus, in this form, the propagator $\mathcal{K}(x,\tau;x_0,\tau_0)$ is of little practical use.

To overcome these difficulties, we make use of the imaginary time *evolution* operator, $e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)}$. When operating on an arbitrary state $\Psi(x,\tau_0)$ at time τ_0, the evolution operator translates the state forward in time by an amount $\Delta\tau$, such that

$$\Psi(x,\tau) = e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)}\Psi(x,\tau_0). \tag{2.44}$$

This can be easily verified by direct application of $e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)}$ to $\Psi(x,\tau_0)$ (equation 2.41 with $\tau = \tau_0$). The utility of the imaginary time evolution operator results from its close association with the propagator $\mathcal{K}(x,\tau;x_0,\tau_0)$. To better demonstrate the relationship between the two, consider

equation 2.44 written as the integral equation

$$\Psi(x,\tau) = \int_{-\infty}^{\infty} e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)}\delta(x-x_0)\Psi(x_0,\tau_0)\,dx_0. \tag{2.45}$$

In this form, it is evident by equation 2.42 that

$$\mathcal{K}(x,\tau;x_0,\tau_0) = e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)}\delta(x-x_0) \tag{2.46}$$

and that equation 2.43 is recovered by application of the closure relation to $\delta(x-x_0)$. Consequently, we can choose to work with the imaginary time evolution operator in place of the propagator.

To accomplish this, we must contend with the power series expansion

$$e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)} = e^{\Delta\tau E_R}\sum_{n=0}^{\infty}\frac{(-\Delta\tau)^n}{n!}\hat{\mathcal{H}}^n \tag{2.47}$$

of the imaginary time evolution operator in terms of increasing powers of the Hamiltonian operator. Choosing a small value for the time increment $\Delta\tau$ has two advantages. First, small values of $\Delta\tau$ permit us to neglect higher order terms in $\hat{\mathcal{H}}$, making the evolution operator more convenient to evaluate. Second, we may approximate the evolution operator as the product

$$e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)} \approx e^{-\Delta\tau\hat{T}}e^{-\Delta\tau\hat{V}}e^{\Delta\tau E_R}, \tag{2.48}$$

separating contributions from the kinetic energy ($\hat{T}$) and potential energy ($\hat{V}$) operators. Most generally, the Campbell-Baker-Haussdorff (CBH) formula

$$e^{-\Delta\tau(\hat{T}+\hat{V})} = e^{-\Delta\tau\hat{T}}e^{-\Delta\tau\hat{V}} - \frac{\Delta\tau}{2}[\hat{T},\hat{V}] - \frac{(\Delta\tau)^2}{12}[\hat{T}-\hat{V},[\hat{T},\hat{V}]] + \ldots \tag{2.49}$$

includes commutators of $\hat{T}$ and $\hat{V}$ which we may neglect for small enough $\Delta\tau$. This is known as the short-time approximation (STA). Equation 2.48 is known as the Trotter formula and is only accurate to $\mathcal{O}((\Delta\tau)^2)$. In contrast, the more commonly used symmetrized version of the Trotter formula

$$e^{-\Delta\tau(\hat{\mathcal{H}}-E_R)} \approx e^{-\Delta\tau\hat{V}/2}e^{-\Delta\tau\hat{T}}e^{-\Delta\tau\hat{V}/2}e^{\Delta\tau E_R}. \tag{2.50}$$

is accurate to third order in $\Delta\tau$. Fourth order expansions also exist, further reducing the error relative to the CBH expansion (equation 2.49).

C. Green's functions and path integrals

The utility of partitioning the evolution operator into both kinetic and potential terms is elucidated by their corresponding physical interpretation. The operator $e^{-\Delta\tau\hat{T}}$ is the evolution operator for a purely diffusive system with no potential present. The Green's function solution for

free diffusion (in k-dimensional space) is given by

$$G_T(\mathbf{x}', \mathbf{x}; \Delta\tau) = \left(\frac{m}{2\pi\hbar^2\Delta\tau}\right)^{k/2} \exp\left[-\frac{m(\mathbf{x}' - \mathbf{x})^2}{2\hbar^2\Delta\tau}\right]. \tag{2.51}$$

The Green's function is an influence function which gives the contribution at $\mathbf{x}'$ arising from the source point $\mathbf{x}$ as a function of time. At $\tau = 0$, the Green's function reduces to a delta function located at $\mathbf{x}' = \mathbf{x}$. This delta function evolves into a Gaussian function, widening with time. Systems with a non-zero potential will also experience this diffusive behavior. Information about the system-specific potential energy is contained within the operator $e^{-\Delta\tau(\hat{V} - E_R)}$. This piece, which is a local operator, is often referred to as the *branching factor* and is denoted $G_B(\mathbf{x}', \Delta\tau)$. Together, the combined short-time Green's function approximation, G_{STA}, is written as the product

$$G_{STA}(\mathbf{x}', \mathbf{x}; \Delta\tau) = G_T(\mathbf{x}', \mathbf{x}; \Delta\tau) G_B(\mathbf{x}', \Delta\tau). \tag{2.52}$$

For small enough values of $\Delta\tau$, G_{STA} is a good approximation to the true Green's function and is exact in the limit of vanishing potential. For a free particle, G_{STA} is exact for at all values of $\Delta\tau$.

For long times, G_{STA} is clearly inadequate as a propagator since it is only valid for small time *increments* (time steps). To accurately represent $\mathcal{K}(\mathbf{x}', \tau; \mathbf{x}, \tau_0)$ in the short-time approximation, it is necessary to carry out many sequential short-time propagations. We may do this using a path integral, treating the kernel as a product of short-time propagators integrated over all intermediate coordinates,

$$\mathcal{K}(\mathbf{x}_S, \tau; \mathbf{x}_0, \tau_0) \approx \int_{-\infty}^{\infty} \left(\prod_{s=1}^{S-1} d\mathbf{x}_s\right) \prod_{s=1}^{S} G_{STA}(\mathbf{x}_s, \mathbf{x}_{s-1}; \Delta\tau), \tag{2.53}$$

where $\Delta\tau$ is equal to the time increment $(\tau - \tau_0)/S$. In the limit $S \to \infty$ this expansion of $\mathcal{K}(\mathbf{x}_S, \tau; \mathbf{x}_0, \tau_0)$ becomes exact. The practical implementation of equation 2.53 is possible using small time increments and a large number of short-time propagations. However, it should be noted that for any finite $\Delta\tau$, the approximate kernel is not equal to the true kernel. Consequently any practical DMC simulation will suffer from a *finite time step bias*, and will fail to generate the exact ground state wavefunction.

D. Stochastic evolution and Markov chains

For an arbitrary initial state (equation 2.41), propagation to large times implies that all excited states have decayed within a threshold of accuracy relative to the ground state. Many short-time propagations are necessary to achieve these long total times. Consequently, the path integral (equation 2.53) will contain an exceedingly large number of intermediate integrals that must be carried out. For such integrals, only stochastic methods of evaluation, such as Monte Carlo, are practical.

We can also relate the path integral itself to a Markov chain, in which each intermediate short-time propagation yields a new state in the chain. This approach forces us to interpret

$G_{STA}(\mathbf{x}', \mathbf{x}; \Delta\tau)$ as a transition probability or matrix element connecting the coordinates $\mathbf{x}'$ and $\mathbf{x}$. A set of such transition probabilities comprises the stochastic propagation matrix, $\mathbf{M}$. A stochastic matrix must by definition be non-negative and normalized to conserve probability; in general, G_{STA} is not normalized in this way. To circumvent this problem, we factor $\mathbf{M}$ into a normalized transition matrix $\mathbf{P}$ and a weight matrix or vector $\mathbf{W}$. $\mathbf{P}$ is constructed from G_T, which also depends on both $\mathbf{x}$ and $\mathbf{x}'$; the entries of $\mathbf{W}$ come from G_B, which depends only on $\mathbf{x}$. Together, $\mathbf{P}$ and $\mathbf{W}$ are used to construct a normalized Markov chain representation of the path integral (equation 2.53). However, such a scheme defeats the generalized Metropolis approach for controlling the time step bias.

An additional consequence of employing Markov chains and stochastic methods for propagation is that we must also choose a stochastic representation of $\Psi(\mathbf{x}, \tau)$. Typically in DMC methods, we represent $\Psi(\mathbf{x}, \tau)$ as a population or ensemble of delta functions whose locations change as we propagate. This representation of $\Psi(\mathbf{x}, \tau)$ imposes a discretization or *finite sampling error* since we are attempting to describe an inherently continuous function with a set of discrete points. The characterization of this error and its contribution to various aspects of a DMC calculation are addressed further in chapter 3.

E. Reference energy and population control

In equation 2.40, we introduced the reference energy, E_R as a device for stabilizing the ground state with respect to the decay of excited states. To ensure a static ground state limit as $\tau \to \infty$, it is necessary to choose $E_R = E_0$; unfortunately, E_0 is initially unknown. We also note that E_R appears in the definition of the branching factor G_B and the weight vector $\mathbf{W}$. Though E_R is unknown, it is still responsible for the overall normalization of the stochastic propagator. If we arbitrarily choose a value for E_R, the propagator will be improperly normalized and the norm of the weight vector will increase or decrease with time (for $E_R > E_0$ or $E_R < E_0$, respectively). Only when E_R is exactly equal to E_0 will the stochastic propagator be properly normalized. However, for short-time propagators the reference energy will not be exactly equal to E_0 since finite time propagations will not generate a pure ground state.

Typically in DMC simulations, E_R is periodically estimated from changes in the population size (which is equivalent to the norm of the weight vector for a delta function population). This estimate of E_R is called the *growth energy*, E_G and is given by the *growth estimator*

$$E_G = \langle E_L \rangle - \frac{\gamma}{\Delta\tau} \ln \left\{ \frac{|\mathbf{W}_s|}{|\mathbf{W}_0|} \right\} \tag{2.54}$$

where γ is an adjustable parameter and $\langle E_L \rangle$ is an estimate of the local energy $\Psi^{-1}\hat{\mathcal{H}}\Psi$. $|\mathbf{W}_0|$ is the initial wavefunction norm and $|\mathbf{W}_s|$ is the wavefunction norm after s propagations. Using E_G to control E_R through periodic adjustments results in a clever "self-consistent" approach to determining E_0 as well as the ground state eigenfunction. However, this method of controlling the reference energy introduces unwanted feedback into the stochastic evolution of the system. Since information from the history of the simulation influences its future state, the stochastic evolution

can no longer be considered a true Markov chain. This introduces an additional *population control bias* and will cause simulation estimates of E_0 to be greater than the true value.[44] In practice, the degree of non-Markovian contamination depends on the size of the delta function population approximating $\Psi(\mathbf{x}, \tau)$ and the frequency with which the reference energy is updated. To reduce the the feedback bias, the reference energy is usually updated only after several propagation steps. This helps to reduce correlation and is the method that we have chosen to use in this work.

2.5.3 Pure diffusion Monte Carlo

The pure diffusion Monte Carlo algorithm is primarily based on a stochastic interpretation of equation 2.53, with the diffusion portion G_T of the short-time propagator G_{STA} (equation 2.51) simulated as a random walk process.

To begin, we select the initial state, $\Psi(\mathbf{x}, \tau_0)$ to be represented by an ensemble of N uniformly weighted delta functions distributed arbitrarily in configuration space. By configuration, we mean that each delta-function represents a single geometry or spatial arrangement of n particles (nuclei, electrons, etc) in the system and is uniquely defined by a set of $k = 3n$ spacial coordinates (typically Cartesian). In the diffusion stage of the short-time propagation, we displace each delta function (or *walker*) in configuration space according to some random vector. For Cartesian coordinates, the components x_i of this vector are random deviates (ξ_i) drawn from a one dimensional standard normal distribution and scaled by a length factor. Coordinates are updated after a move according to

$$x_i' = x_i + \sqrt{2D\Delta\tau}\,\xi_i. \tag{2.55}$$

Next, we apply the branching portion G_B of the short-time propagator to the population, using the potential $V(\mathbf{x}')$ and the current estimate of the reference energy E_R to statistically reweight each delta function according to

$$w' = w\, e^{-\Delta\tau(V(\mathbf{x}')-E_R)} \tag{2.56}$$

where w' is the new weight of a walker at $\mathbf{x}'$ and w is the weight of the same walker at the prior position $\mathbf{x}$. Symmetric splitting (equation 2.50) is usually implemented as

$$w' = w\, e^{-\Delta\tau\left(\frac{1}{2}\left[V(\mathbf{x}')+V(\mathbf{x})\right]-E_R\right)}, \tag{2.57}$$

and used instead of equation 2.56 to reduce the statistical error.

This reweighting is sufficient to maintain the correct overall normalization of the stochastic propagation matrix but is rather computationally inefficient in practice. To make the algorithm more efficient, we would like to keep the weights comparable so that single configurations do not dominate and that insignificant configurations do not monopolize the calculation. We accomplish this by implementing a birth/death process. After the reweighting, a walker whose weight is sufficiently large is split into n copies each with weight $1/n$. Walkers whose weights fall below a threshold cutoff are eliminated. This process reduces statistical fluctuations and is more computationally stable than simple reweighting. Unfortunately, the raw population size (actual number of delta functions) is

free to grow or shrink. Algorithmically, this approach is somewhat trickier to implement, leading some to develop alternate reweighting schemes.[45–48] However, in this work we have chosen to use the birth/death algorithm to moderate the population weights. For sufficiently accurate wavefunctions, such as those employed in this work, weights grow only very slowly with time yielding rather modest growth in the population size.

After the entire population has been updated, population control may be applied and the reference energy updated as necessary. These steps are repeated until the population size has stabilized and the reference energy has reached self-consistency, signaling that the stationary distribution has been practically attained. Once this occurs, properties may be computed over the stationary distribution. Below is sample pseudocode for the pure diffusion Monte Carlo algorithm.

1. Initialize the walker population.
2. Diffuse each walker in the ensemble to a new location in configuration space (equation 2.55).
3. Compute the potential energies at the new positions.
4. Reweight each walker using the branching factor (equation 2.57).
5. Compute/accumulate observables (once the population has equilibrated to the required accuracy).
6. Apply the birth/death process.
7. Update the reference energy (as necessary) and advance time.
8. Repeat steps 2-7 many times.
9. Compute ensemble/simulation averages of observables.

2.5.4 Importance sampled diffusion Monte Carlo

As with VMC, we can design a DMC process having a probability density as the stationary distribution. This stationary distribution evolves in imaginary time according to the Fokker-Planck equation

$$\frac{\partial \rho(\mathbf{x},\tau)}{\partial \tau} = D\,\nabla^2 \rho(\mathbf{x},\tau) - D\,\nabla \cdot \left[\rho(\mathbf{x},\tau)\nabla \ln \Psi_T^2(\mathbf{x})\right] - \left[\frac{\hat{\mathcal{H}}\Psi_T(\mathbf{x})}{\Psi_T(\mathbf{x})} - E_R\right]\rho(\mathbf{x},\tau), \tag{2.58}$$

which yields a mixed probability density $\rho(\mathbf{x},\infty) = \Psi(\mathbf{x})\Psi_T(\mathbf{x})$, instead of $\rho(\mathbf{x},\infty) = \Psi_T^2(\mathbf{x})$ as found in VMC. As before, D is the diffusion constant $\hbar^2/2m$, and $D\nabla \ln \Psi_T^2(\mathbf{x})$ is the quantum force, $\mathbf{F}(\mathbf{x})$. The quantity $\Psi_T^{-1}(\mathbf{x})\hat{\mathcal{H}}\Psi_T(\mathbf{x})$ is referred to as the *local energy*, $E_L(\mathbf{x})$. This equation is similar to the Schrödinger equation (equation 2.35) with the exception of an additional gradient term which imposes a "drift" on the diffusion process. This is more clearly seen in the short time

Green's function solution to the Fokker-Planck equation

$$G_T^{FP}(\mathbf{x}', \mathbf{x}; \Delta\tau) = \left(\frac{m}{4\pi D\Delta\tau}\right)^{k/2} \exp\left[-\frac{(\mathbf{x}' - \mathbf{x} - D\Delta\tau\mathbf{F}(\mathbf{x}))^2}{4D\Delta\tau}\right] \tag{2.59}$$

which is similar to G_T (equation 2.51) but with the quantum force now appearing in the exponential. We implement this extra term with the modified random walk

$$\mathbf{x}' = \mathbf{x} + D\Delta\tau\mathbf{F}(\mathbf{x}) + \sqrt{2D\Delta\tau}\,\vec{\xi} \tag{2.60}$$

where $\vec{\xi}$ is a normally distributed random vector. This modified diffusion process is identical to that of the Fokker-Planck implementation (equation 2.21) of variational Monte Carlo, and by itself would generate the stationary distribution Ψ_T^2. This means that the third term on the right hand side of equation 2.58 must be responsible for generating the mixed distribution $\Psi\Psi_T$ from Ψ_T^2. This occurs through the application of a modified branching factor

$$e^{-\Delta\tau\left(\frac{1}{2}\left[E_L(\mathbf{x})+E_L(\mathbf{x}')\right]-E_R\right)}, \tag{2.61}$$

which effectively introduces a multiplicative factor of Ψ/Ψ_T into the distribution Ψ_T^2 to give the mixed distribution. It should be noted that in the limit $\Psi_T = 1$, the branching factor is solely responsible for generating Ψ and we recover the pure DMC algorithm. As Ψ_T approaches Ψ, the branching factor will contribute nothing and the burden of generating Ψ^2 will rest completely on the modified diffusion process. Since Ψ_T is arbitrary, we can take advantage of the "tunable" nature of this *importance sampling* algorithm. Choosing a trial function, Ψ_T, close to the true form of Ψ improves the efficiency of the method and decreases statistical fluctuations by reducing the influence of the branching factor.

As was the case with the VMC implementation of the Fokker-Planck algorithm, the guided random walk approach is only correct in the limit $\Delta\tau \to 0$. This means that the importance sampled DMC algorithm will also suffer from the effects of finite time step bias. Unfortunately, the generalized Metropolis procedure of accepting or rejecting moves according to

$$A(\mathbf{x} \to \mathbf{x}') = \min\left\{1, \frac{|\Psi_T(\mathbf{x}')|^2 G_T^{FP}(\mathbf{x}', \mathbf{x}; \Delta\tau)}{|\Psi_T(\mathbf{x})|^2 G_T^{FP}(\mathbf{x}, \mathbf{x}'; \Delta\tau)}\right\} \tag{2.62}$$

does not compensate for the time step bias as it did in the corresponding VMC case. Here, it merely guarantees that we evolve the correct stationary distribution as $\Psi_T \to \Psi_0$. This is because of the incorporation of the additional branching factor into the stochastic evolution to preserve the normalization. Consequently, an extrapolation of $\Delta\tau \to 0$ is the only means of removing this bias from the DMC algorithm. Finite time step bias can also be reduced through the use of higher order propagators. As with VMC, we must also work to suppress the correlation error within DMC by controlling the rejection rate and avoiding persistent configurations. As before, delayed rejection is an effective method of achieving this.

The importance-sampled algorithm is very similar to the pure DMC algorithm but has a few important differences. First, the diffusive random walk has been replaced by a guided random walk. Second, the local energy is used instead of the potential energy for computing the branching factor. Third, we have added a generalized Metropolis acceptance step with delayed rejection to the move. These relatively minor changes provide a significant improvement in the efficiency of the DMC algorithm and substantially reduce statistical fluctuations. Pseudocode for a representative importance-sampled DMC algorithm appears below.

1. Initialize the walker population.
2. Diffuse each walker in the ensemble to a new location in configuration space (equation 2.60).
3. Compute the local energies at the new positions.
4. Accept or reject the move based on the generalized Metropolis scheme (equation 2.62).
5. If the move is accepted, reweight each walker using the branching factor (equation 2.61).
6. If the move is rejected, retain the original weight and local energy or attempt a delayed rejection move.
7. Compute/accumulate descendant weighted observables (if the population is equilibrated).
8. Apply the birth/death process.
9. Update the reference energy (as necessary) and advance time.
10. Repeat steps 2-9 many times.
11. Compute ensemble/simulation averages of observables.

2.5.5 Computing observables in DMC

A. Extrapolation

In the PDMC algorithm, we generate a stationary distribution which is proportional to the ground state wavefunction. This complicates the computation of quantum mechanical observables because we do not have direct access to the probability density Ψ^2. Unfortunately, importance sampling does not solve this problem, as it yields the mixed probability density $\Psi\Psi_T$ instead of the correct quantum mechanical density Ψ^2. Consequently, estimates of any property will depend directly on the quality of the trial wavefunction.

To obtain accurate and unbiased estimates of observables, we would prefer to compute the *pure estimate*

$$\langle A \rangle_p = \frac{\langle \Psi \mid \hat{A} \mid \Psi \rangle}{\langle \Psi \mid \Psi \rangle} \tag{2.63}$$

for an arbitrary coordinates or local operator $\hat{A}$ instead of the *mixed estimate*

$$\langle A \rangle_m = \frac{\langle \Psi_T \,|\, \hat{A} \,|\, \Psi \rangle}{\langle \Psi_T \,|\, \Psi \rangle} \tag{2.64}$$

obtained from the importance sampled algorithm. A commonly used method for estimating $\langle A \rangle_p$ is to incorporate information from both the mixed and variational estimates to obtain an *extrapolation estimate* of this quantity. Taking $\delta\Psi = \Psi - \Psi_T$ to be the difference between the exact wavefunction and the approximate trial wavefunction, we may employ the expected value $\langle \delta\Psi \,|\, \hat{A} \,|\, \delta\Psi \rangle$ to easily develop the relation

$$\langle A \rangle_e = 2\langle A \rangle_m - \langle A \rangle_v = \langle A \rangle_p + \mathcal{O}(\delta\Psi^2). \tag{2.65}$$

Here, the extrapolation estimate $\langle A \rangle_e$ is an approximation of $\langle A \rangle_p$ which is accurate at second order in $\delta\Psi$. The utility of this estimator lies in the readily available quantities $\langle A \rangle_m$ and $\langle A \rangle_v$; no modifications to either the VMC or ISDMC algorithms are required. Unfortunately, the accuracy of this expression still depends on the quality of the trial wavefunction, in addition to the statistical uncertainties and biases present in the computed estimates $\langle A \rangle_m$ and $\langle A \rangle_v$.

B. Forward walking/descendant weighting

Instead of extrapolating, we can modify the importance sampling algorithm to sample observables from the true probability density. This is accomplished by the introduction of an additional factor of Ψ/Ψ_T into the reweighting process.[49] Instead of reweighting the distribution, we may reweight the contribution of each observable $\hat{A}$ to give the estimate

$$\langle A \rangle_{FW} = \frac{\langle \Psi_T \,|\, (\Psi/\Psi_T)\hat{A} \,|\, \Psi \rangle}{\langle \Psi_T \,|\, (\Psi/\Psi_T) \,|\, \Psi \rangle}. \tag{2.66}$$

The trick comes in computing Ψ/Ψ_T, which arises naturally from application of the branching factor. At long times, Ψ/Ψ_T is estimated from the total weight of all walkers that are descended from a single walker. *Forward-walking* algorithms tag and track descendants[45,50,51] in an effort to estimate this ratio. Unfortunately, forward walking algorithms are considered largely unstable[46,52,53] and have been shown to be unreliable in certain calculations. Casulleras and Boronat have developed a complementary approach[46,54] wherein statistically weighted observables are inherited and modified by descendants. This implementation of forward walking is commonly referred to as *descendant weighting* and has been successfully employed in the study of cryogenic atomic clusters[27] and bulk superfluids.[55] We expect the descendant weighting scheme to work well for the present investigation of van der Waals clusters, since this problem is similar to previous successful applications and our potentials lack problematic features such as a Coulombic singularity.

In the descendant weighting method, weights are evolved for a period of time by a combi-

nation of reweighting and accumulation of observables

$$a' = w' A(\mathbf{x}) + \left\{ \frac{w'}{w}\, a \right\} \tag{2.67}$$

followed by a period of relaxation

$$a' = \left\{ \frac{w'}{w} a \right\} \tag{2.68}$$

in which the weights approach their asymptotic values. Generally, the relaxation period must be long enough for all excited state components to decay out of the last accumulated observation. Descendant weights are also subject to correlations and must be sampled from independent or decorrelated configurations to avoid a correlation error in estimates of the observable $\hat{A}$. We have elected to use the modified approach of Boronat and Casulleras for the current study.

2.6 Summary of sources of bias in QMC methods

2.6.1 Correlation bias

With the exception of perhaps the basic VMC Metropolis algorithm, all other VMC and DMC methods suffer from some form of correlation bias. This bias is introduced when sequential samplings fail to yield fully independent estimates of properties. Sampling configurations can become correlated in time as a result of small step sizes or because of excessive or persistent rejections of attempted moves. The removal of this bias requires blocking procedures that group sequential estimates into single, independent snapshot estimates. A simple autocorrelation analysis of the Monte Carlo data provides the correlation time and block sizes necessary to sufficiently eliminate this source of bias.

2.6.2 Non-equilibrium bias

This bias can occur if the random walk process is unable to sample all regions of configuration space during the course of the simulation. In such a case, the associated Markov chain fails to be effectively ergodic and will not yield the desired stationary distribution. This could occur in situations in which multiple potential minima are separated by formidable barriers. For very weakly bound systems like He_2 close to the dissociation limit, the equilibration time is lengthened because of the long times necessary for neighboring continuum states to decay. To suppress the non-equilibrium bias, we typically throw away an initial period of the random walk. Observables may then be computed over the remainder of the simulation. Initially dispersing sampling points throughout configuration space can also help to mitigate this problem and at the same time accelerate convergence.

2.6.3 Finite time step bias

All QMC methods that employ a short-time propagator are susceptible to time step bias. This is a consequence of the short time propagator being an approximation of the true Green's function propagator. Since we never know the true propagator for a realistic system, we must partition the total time evolution of the system into many short time slices. As the number of time slices becomes infinite and the time step approaches zero, the approximation becomes exact. For all other finite values of the time step, the steady state distribution will not be equal to the exact one. In the Fokker-Planck VMC approach, we can correct this problem by enforcing detailed balance in a generalized Metropolis scheme. For DMC, the partitioning of the stochastic propagator into a normalized transition matrix and a weight vector defeats our ability to correct the finite time evolution with a detailed-balance step. Thus for DMC, extrapolation of quantities to zero time step is required to suppress the time step bias and is the method we have chosen to use in the present study.

The time step error also manifests itself differently with respect to the local energy and growth energy (equation 2.54) estimators. It has been shown[44] that the growth energy is biased lower than the average local energy by an amount linear in the size of the time step:

$$\langle E_G \rangle - \langle E_L \rangle = -\frac{\Delta\tau}{2M} \sum_{m=1}^{M} \left(\langle E_L^2 \rangle_{m-1} - \langle E_L \rangle_{m-1}^2 \right) + O(\Delta\tau^2) \tag{2.69}$$

Here, the subscripts $m-1$ refer to averages over the previous snapshot. The difference between expectation values of the local energy and growth energy estimators is proportional to the size of the time step, and decreases to zero as $\tau \to 0$. Thus, the discrepancy between these energy estimates at a finite value of $\Delta\tau$ is a useful qualitative indicator of the level of time step bias present in the simulation. In general, the slope of the linear dependence will change depending on the quality of the wavefunction.

2.6.4 Population control bias

Population control or feedback bias is unique to DMC and is a result of adjusting the reference energy to keep the Markov chain normalized. Population control uses information about the history of the system to control the time evolution of that system. In this respect, the Markov property is violated and the associated random walk results in a slightly different stationary distribution. For small populations, the degree of non-Markovian contamination can be considerable.

To completely eliminate the population control bias, feedback should not be used. DMC simulations in the absence of population control are sometimes practical given a very accurate trial wavefunctions and a good prior estimate of the true energy. In situations where population control must be used to stabilize the algorithm, adjustments to the reference energy should take place as infrequently as possible. This minimizes the introduction of non-Markovian information into the stochastic evolution. As with other forms of bias, extrapolation techniques can also be employed to

obtain values of properties in the limit of no population control.

2.6.5 Finite population bias of the growth estimator

In DMC, the growth energy is also subject to a finite population bias. This bias is related to the covariance between fluctuations in the local energy and fluctuations in the finite population weights[44]

$$\langle E_G \rangle - \langle E_L \rangle = -\frac{1}{M} \sum_{m=1}^{M} \frac{1}{\langle |\mathbf{W}| \rangle_m} \left(|\mathbf{W}_m| - \langle |\mathbf{W}| \rangle_m \right) \left(\langle E_L \rangle_m - \langle E_L \rangle \right) \tag{2.70}$$

Here we have assumed $\Delta\tau \to 0$ so that the time step error can be neglected. Since the fluctuations in the local energy and in the weights are of opposite sign, the net effect of the finite population bias is to cause the growth energy estimate to be higher than that of the local energy estimate. This bias is inversely proportional to the norm of the population and is present even when the reference energy is not updated.

2.6.6 Finite population bias of descendant weights

In the Boronat and Casulleras descendant weighting scheme,[46,54] observables are inherited and reweighted by descendants according to the ratio w'/w (equation 2.68). Expectation values of descendant weighted properties are then averages over these ratios. Unfortunately, for a finite sampling population, the expectation value of a ratio is not equal to a ratio of expected values. This means that descendant weighted properties will exhibit a bias for all realistic population sizes that will generally be more pronounced in higher dimensional systems. Currently, the only method of dealing with this bias is by extrapolation to infinite population. In the case of descendant-weighted distribution functions, extrapolation to infinite population is computationally prohibitive. Thus, in the present study, we report unextrapolated descendant-weighted distribution functions.

2.6.7 Relaxation bias of descendant weights

A singe descendant weight is collected at one configuration and time in the simulation. Over time, as the walker visits other configurations drawn from the stationary distribution, the descendant weight will shrink or grow and be passed on when the walker is split. This means that to obtain a good estimate of Ψ_0/Ψ_T, the walker must visit many configurations and allow the descendant weight estimate to equilibrate before an expectation value is taken. This equilibration can be compared to the initial simulation equilibration of a population of walkers that all start out at a single point. For small time steps, this equilibration can be lengthy. However, unlike population equilibration, each time we sample an observable (equation 2.67), the equilibration period starts anew. Therefore, a lengthy equilibration or relaxation period is required after final observable sampling to ensure that the descendant weights have converged. If this relaxation period is neglected, a bias will be introduced into descendant weight estimates.

Chapter 3

Dimensionality and Finite Sampling Errors in QMC

3.1 Overview

In this chapter, a theoretical approach to the analysis of statistical uncertainties in QMC observables is presented. This method examines the dependence of the statistical error on basic factors such as dimensionality, importance sampling, population size, and number of population snapshots. The approach begins with an examination of the stationary distribution and the extraction of meaningful statistical information from the stationary distribution through the use of projectors. From there, functionals of these projectors are introduced for the calculation of the population and sample variances and for computing the variation or error in these variances. Statistical distributions of the coefficient operator and arbitrary coordinate operators, as well as their corresponding variance distributions are discussed. This chapter provides a theoretical underpinning for the work presented in later chapters.

3.2 The stationary distribution

Consider a Monte Carlo sampling of some normalized probability distribution which we will refer to as the stationary distribution, $\pi(\mathbf{x})$. The stationary distribution may have a functionally known form or may be the converged, steady state distribution of a Markov process. Depending on the application, the stationary distribution may be proportional to a wavefunction (as in standard DMC) or to a probability density (as in importance-sampled DMC). In this work, the stationary distribution is always taken to be a nodeless, bosonic ground state wavefunction or probability density. This distinction is important since excited state or fermionic wavefunctions (which have negative regions) cannot be normalized as standard probability distribution functions. It should also be noted that $\pi(\mathbf{x})$ is continuous and, in general, a multidimensional function of the state vector $\mathbf{x} = (x_1, x_2, ..., x_k)$.

Monte Carlo sampling restricts our instantaneous knowledge of $\pi(\mathbf{x})$ to a finite number of discrete samples. Accordingly, we define an approximate or empirical stationary distribution, $\tilde{\pi}(\mathbf{x})$, to represent this discrete form of $\pi(\mathbf{x})$. By discrete point-wise sampling, the approximate stationary distribution may be constructed as a normalized sum of delta functions

$$\tilde{\pi}_N(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x}-\mathbf{x}_i) \tag{3.1}$$

where N is the number of point-wise samples taken from $\pi(\mathbf{x})$ and is often referred to as the population size. The larger the number of samples, the better $\tilde{\pi}_N(\mathbf{x})$ approximates $\pi(\mathbf{x})$. As an example, figure 3.1 shows a single thirty delta function approximation to a one-dimensional Gaussian stationary distribution. At first glance, the two distributions appear to have little in common. For comparison, we could histogram the delta functions into bins along x; however, the cumulative (integrated) distribution (figure 3.2) better illustrates the similarity between $\pi(\mathbf{x})$ and the discretized approximation $\tilde{\pi}_N(\mathbf{x})$. The question of how well a sum of N delta functions approximates the true continuous stationary distribution is an important question and will be addressed in this chapter.

As mentioned above, $\pi(\mathbf{x})$ is a multivariate distribution. Unless otherwise stated, all coordinates will be taken to be *independent* or uncoupled. This permits us to write the multidimensional

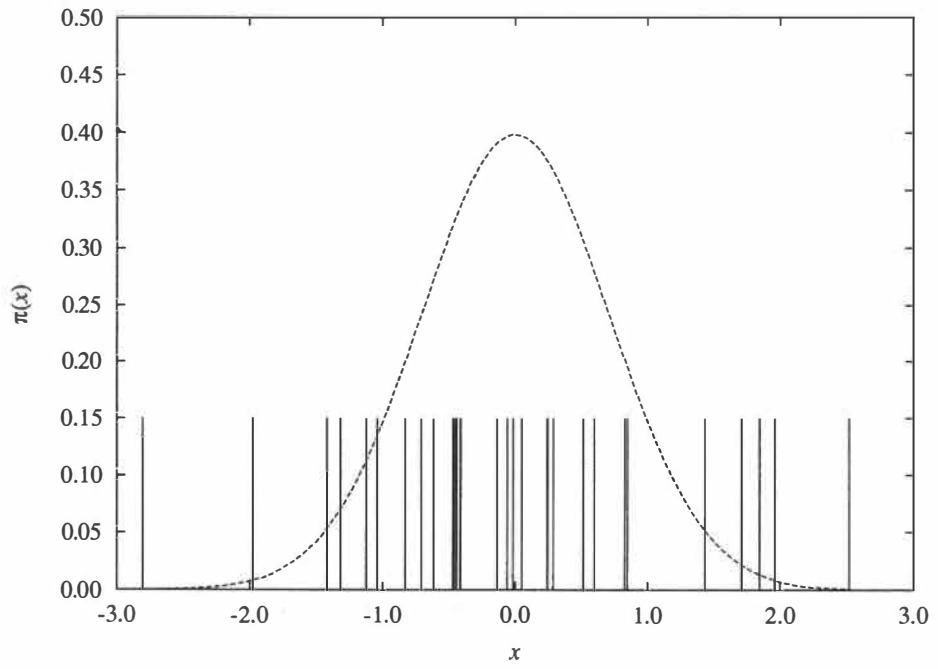


Figure 3.1: An instantaneous thirty δ-function approximation of a Gaussian distribution

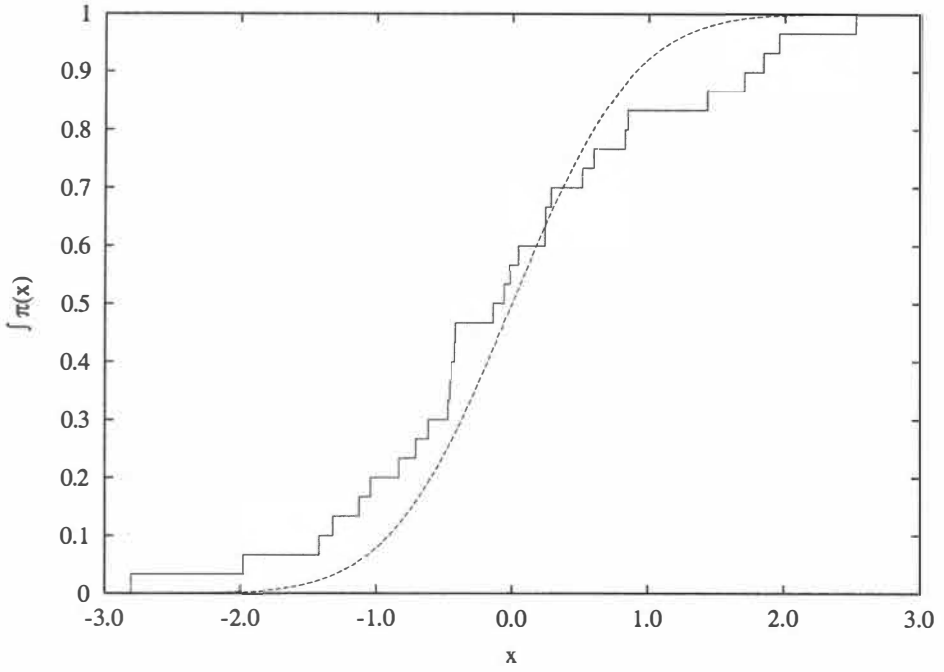


Figure 3.2: A cumulative thirty δ-function approximation of a Gaussian distribution

stationary distribution as a product of one dimensional stationary distributions

$$\pi(\mathbf{x}) = \prod_{i=1}^{k} \pi(x_i) \tag{3.2}$$

where k is the dimensionality or total number of uncoupled coordinates in the system. In most cases, we will also take the stationary distribution to be *isotropic* or identical in all dimensions. These two assumptions will allow us to simply and concisely develop a number of general results. Neither of these assumptions are required; however, they will facilitate this analysis, simplifying a number of equations and clearly demonstrating the dimensional dependence of certain quantities.

3.3 Projectors

Given some approximate (discrete) distribution, $\tilde{\pi}_N(\mathbf{x})$, we seek to find the error due to its "inexactness" when compared with the true continuous stationary distribution, $\pi(\mathbf{x})$. Standard statistical tests such as the Kolmogorov-Smirnov test,[56] the Anderson-Darling test,[57] and the χ^2 test[58] are useful for determining how an approximate distribution mimics another distribution. However, these tests are typically applied to standard one dimensional statistical distributions such as the normal, log-normal, and χ^2 distributions where the quality of these tests is proven and tabulated statistics already exist. Since we are dealing with multivariate distributions that change with the

system or problem of interest, it is reasonable to adopt a more general method of comparison.

It is common in quantum mechanics to use an overlap integral to express the degree of similarity (overlap) between two wavefunctions. For square integrable wavefunctions, the overlap integral takes on some positive value, equal to unity if the wavefunctions are functionally identical and normalized. In the case of orthogonal wavefunctions, the overlap coefficient is exactly zero. We extend the idea of the overlap integral to arbitrary stationary distributions.

Consider the integral

$$c_n = \int f_n(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} \tag{3.3}$$

where we integrate some function $f_n(\mathbf{x})$ against $\pi(\mathbf{x})$ over the support of the stationary distribution. Let $f_n(\mathbf{x})$ be some arbitrary function taken from a complete set of functions, orthogonal over the same support. This defines a prescription for computing an overlap coefficient c_n. If we construct one function, say $f_0(\mathbf{x})$, to be proportional to $\pi(\mathbf{x})$ but suitably renormalized to yield an overlap of unity with $\pi(\mathbf{x})$,

$$1 = \int f_0(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \tag{3.4}$$

we see that $f_0(\mathbf{x})$ becomes a *projector* for the stationary distribution. This is more clearly evident in the case where $\pi(\mathbf{x})$ is replaced with some approximate distribution, such as the distribution $\tilde{\pi}_N(x)$ formed from a population of N delta-functions (equation 3.1). In this case, we see that the integral

$$\tilde{c}_{0,N} = \int f_0(\mathbf{x})\tilde{\pi}_N(\mathbf{x})d\mathbf{x} \tag{3.5}$$

can be interpreted as "projecting" out the part of $\tilde{\pi}_N(\mathbf{x})$ that overlaps with $f_0(\mathbf{x})$. As $N \to \infty$, $\tilde{\pi}_N(\mathbf{x})$ approaches $\pi(\mathbf{x})$ and $\tilde{c}_{0,N}$ tends toward unity, becoming equal to c_0. Expanding $\tilde{\pi}_N(\mathbf{x})$, we see that

$$\tilde{c}_{0,N} = \int f_0(\mathbf{x})\frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x}-\mathbf{x}_i)\,d\mathbf{x} = \frac{1}{N}\sum_{i=1}^{N} f_0(\mathbf{x}_i) \tag{3.6}$$

and that the coefficient $\tilde{c}_{0,N}$ is itself an average of N $\tilde{c}_{0,1}$ coefficients, where each coefficient is given by the projector $f_0(\mathbf{x})$ evaluated at a particular value of $\mathbf{x}_i$ drawn from $\pi(\mathbf{x})$.

Thus, using a single overlap coefficient, $\tilde{c}_{0,N}$, we may compare a discrete population with the corresponding stationary distribution from which it is drawn. Over the course of a Monte Carlo simulation, we draw many such discrete populations (snapshots) from the stationary distribution, each yielding a different coefficient. As the number of snapshots becomes large, the mean value of these coefficients, $\langle\tilde{c}_{0,N}\rangle$ tends to c_0, the exact mean, which is obtained by replacing $\tilde{\pi}_N(\mathbf{x})$ with $\pi(\mathbf{x})$ in equation 3.5. The larger the population, the more $\tilde{\pi}_N(\mathbf{x})$ resembles $\pi(\mathbf{x})$ and the less variation there will be in estimates of the coefficient mean.

In this work, we will refer to projectors using the notation $P_n(\mathbf{x})$ or simply P_n, without showing the explicit dependence on $\mathbf{x}$. The subscript denotes the projector for state n, where $n = 0$ refers to the "ground state" projector and $n > 0$ refers to excited state projectors. This labeling is most convenient when the projectors are directly proportional to the eigenfunctions of the system

of interest. Note that we are not required to use any particular set of projectors as long as the chosen set is integrable over the support of $\pi(\mathbf{x})$. For example, we could project the stationary distribution of the Morse oscillator onto a set of harmonic oscillator eigenfunctions. Labeling $n = 0$ the "ground state" projector is still intuitively useful, even though in this instance, the harmonic oscillator ground state projector is just an approximation to the true Morse ground state projector.

3.4 Expectation values and operator distributions

We can extend this projector to describe matrix elements of *any* operator, not just the coefficient or identity operator, $\hat{1}$. If we consider the ground state matrix element

$$\langle A \rangle = \langle P_0(\mathbf{x}) | \hat{A}(\mathbf{x}) | \pi(\mathbf{x}) \rangle \tag{3.7}$$

of the coordinate operator $\hat{A}(\mathbf{x})$, we recover a standard overlap integral in the case $\hat{A}(\mathbf{x}) = \hat{1}$. If $\hat{A}(\mathbf{x})$ is some other operator, this integral will apply the projector $P_0(\mathbf{x})$ to the function $\hat{A}(\mathbf{x})\pi(\mathbf{x})$. We can instead choose to lump both $P_0(\mathbf{x})$ and $\hat{A}(\mathbf{x})$ together into a new operator $\hat{\Omega}(\mathbf{x})$ and take the matrix element of $\hat{A}(\mathbf{x})$ (equation 3.7) to be the expectation value

$$\langle \Omega \rangle = \langle \hat{\Omega}(\mathbf{x}) | \pi(\mathbf{x}) \rangle \tag{3.8}$$

of $\hat{\Omega}(\mathbf{x})$ over the stationary distribution $\pi(\mathbf{x})$, provided $\hat{\Omega} = P_0 \hat{A}$. For example, the expectation value of the operator $\hat{A} = \mathbf{x}^2$ would be given by

$$\langle \mathbf{x}^2 \rangle = \langle \Omega \rangle = \langle P_0 \mathbf{x}^2 | \pi(\mathbf{x}) \rangle. \tag{3.9}$$

In like manner, the expectation value of any coordinate or local operator may be computed. For the sake of clarity, we use $\hat{\Omega}$ to refer to the product $P_0 \hat{A}$ throughout the remainder of this work. Thus, references to $\hat{\Omega}$ may be replaced by the product of any desired projector and observable without loss of generality. Where appropriate, we will also make use of the abbreviated notation $\hat{\Omega}(\hat{A})$ to disambiguate between omega operators corresponding to different observables.

An advantage of this approach is that the statistical distributions and sampling moments for arbitrary observables are readily accessible. To clarify, let us consider a single observation or random variable $U = \hat{\Omega}(\mathbf{x})$. Unlike the values of $\mathbf{x}$ which are distributed according to $\pi(\mathbf{x})$, observed values of U are distributed according to some new distribution $g(u)$ such that

$$\int \hat{\Omega}(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x} = \int u \, g(u) \, du. \tag{3.10}$$

Values of u are related to estimates of properties. For the choice $\hat{\Omega}(E_L(\mathbf{x}))$, values of u are related to observations of the local energy taken from a distribution of local energies $g(u)$.

This form is particularly convenient for investigating the statistical properties of weighted sums of estimates and is important for characterizing the statistical error of a snapshot estimate

that depends on contributions from each sampling point in the population. We can define a new random variable

$$W = \frac{1}{N}\sum_{i=1}^{N} U_i \tag{3.11}$$

that is an average of N independent observations of U. Values of W are distributed according to the distribution $g(w)$. At an infinitely large population size, each estimate of W becomes equal to the expected value of U and has no statistical variance. In this limit, the distribution $g(w)$ must be characterized as a delta function located at the single value $\langle w \rangle = \langle u \rangle$. For finite sampling populations, the distribution $g(w)$ will always have a smaller variance than $g(u)$. This is in agreement with the theory of sampling statistics which tells us that the variances of identically distributed random samples are related according to

$$\sigma_w^2 = \frac{\sigma_u^2}{N}. \tag{3.12}$$

Thus, we see that the statistical error in values of U or $\hat{\Omega}(\mathbf{x})$ will decrease with increasing population size. Since the error scales as the square root of the variance, this is the origin of the well known $N^{-1/2}$ dependence of the error in Monte Carlo with increasing population size.

3.5 Sampling moments and the statistical error

In the previous section, expectation values have been taken over the exact stationary distribution $\pi(\mathbf{x})$. These expectation values are exact. However, if we were to apply $\hat{\Omega}$ to an approximate distribution $\tilde{\pi}(\mathbf{x})$, the resulting expectation value would only be an estimate of the true expectation value. To differentiate sample averages from true expectation values, we will use the notation $\bar{\Omega}$ to denote a sampling average associated with the approximate distribution $\tilde{\pi}(\mathbf{x})$:

$$\bar{\Omega} = \langle \hat{\Omega} | \tilde{\pi}(\mathbf{x}) \rangle. \tag{3.13}$$

Values of $\bar{\Omega}$ for different snapshots will be denoted with a subscript, $\bar{\Omega}_m$, and simulation averages of $\bar{\Omega}$ will be expressed as $\langle \bar{\Omega} \rangle$.

In general expectation values of an operator are integrated properties of the stationary distribution. In practice, estimates of these expectation values involve summations over sampling points distributed according to the stationary distribution. With every set of sampling points (snapshot) chosen, a new estimate of the expectation value results. Consequently, the estimates will fluctuate from snapshot to snapshot and will have some statistical variance. We introduce the population variance operator

$$\hat{V}_M^2 = \frac{1}{M}\sum_{m=1}^{M} \left(\hat{\Omega}_m - \langle \Omega \rangle\right)^2 \tag{3.14}$$

to quantify this variation in expectation values of the operator $\hat{\Omega}$. Here, M is the number of population snapshots and the subscript m denotes a specific snapshot. The subscript M on the operator $\hat{V}_M^2$ in equation 3.14 indicates that $\hat{V}_M^2$ operates on the expanded state vector $|\pi_1 \pi_2 \ldots \pi_M\rangle$ comprised of M snapshots of the true stationary distribution. Thus, $\hat{V}_M^2$ is integrated according to the multidimensional integral

$$\langle V_M^2 \rangle = \int \hat{V}_M^2(\mathbf{x}) \left[\prod_{m=1}^{M} \pi_m(\mathbf{x}) \right] d^M \mathbf{x} \tag{3.15}$$

and has expectation value

$$\langle V_M^2 \rangle = \langle \hat{V}_M^2 | \pi_1 \pi_2 \ldots \pi_M \rangle = \frac{1}{M} \sum_{m=1}^{M} \langle \Omega^2 \rangle_m - \langle \Omega \rangle^2. \tag{3.16}$$

However, since each copy of π is identical, the sum collapses to yield

$$\langle V^2 \rangle = \langle \Omega^2 \rangle - \langle \Omega \rangle^2 \tag{3.17}$$

where expectation values of Ω are taken over a single copy of the stationary distribution as in equation 3.8 and the M dependence has been dropped. This quantity is the statistical variance of estimates of $\hat{\Omega}$ taken from snapshot to snapshot.

As noted in the preceding section, the variance is also affected by the population size N. Larger sampling populations yield smaller errors, and in the limit of an infinite population the statistical variance must tend toward zero. This dependence is directly obtained from the expectation value of $\hat{V}_{N,M}^2$ which depends on the values of N simultaneous samplings of π_M taken from $\pi_{N,M}$. As noted earlier, these distributions will produce expectation values with variances proportional to N^{-1}. Thus we have

$$\langle V_{N,M}^2 \rangle = \langle \hat{V}_{N,M}^2 | \pi_{N,1} \pi_{N,2} \ldots \pi_{N,M} \rangle = \frac{1}{N} \langle \hat{V}_M^2 | \pi_1 \pi_2 \ldots \pi_M \rangle = \frac{1}{N} \langle V^2 \rangle \tag{3.18}$$

where the distributions π_M are identical to $\pi_{1,M}$. This gives us the statistical variation in expectation values of $\hat{\Omega}$ computed over M snapshots, each sampled with a population of size N.

We can obtain *estimates* of the statistical variance when we perform a similar analysis with $\tilde{\pi}_N$ substituted for π_N. For comparison with actual simulation data, we must use the sample variance operator

$$\hat{S}_{N,M}^2 = \frac{1}{M-1} \sum_{i=1}^{M} \left(\bar{\Omega}_m - \langle \bar{\Omega} \rangle \right)^2 \tag{3.19}$$

which is an unbiased estimator for the population variance, $\langle V_{N,M}^2 \rangle$. The factor of $M-1$ takes into account the loss of a degree of freedom when estimating the mean from the data. Expectation values of $\hat{S}_{N,M}^2$ are computed over the state vector $|\tilde{\pi}_{N,1} \tilde{\pi}_{N,2} \ldots \tilde{\pi}_{N,M}\rangle$, where all snapshots $(\tilde{\pi}_{N,M})$ of the approximate stationary distribution are identically distributed (but not identical) sums of N

δ-functions. These δ-function sums will differ from snapshot to snapshot and are drawn from the same distribution, $\pi(\mathbf{x})$.

Like expectation values of the $\hat{V}^2_{N,M}$ operator, $\hat{S}^2_{N,M}$ should also depend on N. Recall that in the case of $\hat{V}^2_{N,M}$, it is necessary to operate on the state vector $|\pi_{N,1}\pi_{N,2}\ldots\pi_{N,M}\rangle$ composed of probability distributions that differ with the population size N. As a sum of delta functions, $\tilde{\pi}_N$ already depends on the population size N and will intrinsically produce variances proportional to N^{-1}. Additionally, the expectation value $\bar{S}^2_{N,M}$ will be slightly larger (by a factor of $(M/(M-1))$ than the corresponding exact population variance given by $\langle V^2_{N,M}\rangle$.

As an aside, the statistical variance of an operator $\hat{\Omega}$ obtained from expectation values of $\hat{V}^2_{N,M}$ or $\hat{S}^2_{N,M}$ should not be confused with the quantum mechanical variance of an observable. The quantum mechanical variance of some operator $\hat{A}$ is given by the expectation value $\langle\Omega(\hat{A}^2 - \langle A\rangle^2)\rangle$ (equation 3.8). In the same manner, $\hat{V}^2_{N,M}$ or $\hat{S}^2_{N,M}$ may be used to compute the statistical variance of the quantity $\hat{A}^2 - \langle A\rangle^2$. The statistical variance arises from the discrete sampling of an inherently continuous distribution. Real observables have no such sampling error.

Since expectation values of $\hat{S}^2_{N,M}$ are only *estimates* of the true variance given by $\langle V^2_{N,M}\rangle$, we expect $\hat{S}^2_{N,M}$ to also be subject to statistical fluctuations. This variance of the variance is described by the operators

$$\hat{\mathbf{V}}^2_L = \frac{1}{L}\sum_{\ell=1}^{L}\left(\hat{V}^2_{N,M,\ell} - \langle V^2_{N,M,L}\rangle\right)^2 \tag{3.20}$$

and

$$\hat{\mathbf{S}}^2_L = \frac{1}{L-1}\sum_{\ell=1}^{L}\left(\bar{S}^2_{N,M,\ell} - \langle \bar{S}^2_{N,M,L}\rangle\right)^2 \tag{3.21}$$

taken over L sets or simulations, each comprised of M snapshots. Like the $\hat{V}^2_{N,M}$ operator, $\hat{\mathbf{V}}^2_L$ can also be reduced to give (proof A.4)

$$\langle \mathbf{V}^2_L\rangle = \frac{1}{MN^3}\Big(\langle\Omega^4\rangle - 4\langle\Omega^3\rangle\langle\Omega\rangle + (2N-3)\langle\Omega^2\rangle^2 - 4(N-3)\langle\Omega^2\rangle\langle\Omega\rangle^2 + 2(N-3)\langle\Omega\rangle^4\Big) \tag{3.22}$$

in terms of one dimensional integrals or raw moments of a single stationary distribution. Similarly, the expectation value of $\hat{\mathbf{S}}^2_L$ can also be related to moments of the stationary distribution and is found to be (proof A.5)

$$\begin{aligned}\langle \hat{\mathbf{S}}^2_L\rangle \;=\; & \left(\frac{L}{L-1}\right)\frac{1}{MN^3}\left\{\langle\Omega^4\rangle - 4\langle\Omega^3\rangle\langle\Omega\rangle + \left(\frac{2NM-3M+3}{M-1}\right)\langle\Omega^2\rangle^2\right.\\ & \left. -4\left(\frac{NM-3M+3}{M-1}\right)\langle\Omega^2\rangle\langle\Omega\rangle^2 + 2\left(\frac{NM-3M+3}{M-1}\right)\langle\Omega\rangle^4\right\},\end{aligned} \tag{3.23}$$

which for large values of M and L reduces to $\langle \mathbf{V}^2_L\rangle$ or equation 3.22. Unlike expectation values of $\hat{S}^2_{N,M}$ and $\hat{V}^2_{N,M}$ discussed above, which are related by the simple expression

$$\langle \bar{S}^2_{N,M}\rangle = \frac{M}{M-1}\langle V^2_{N,M}\rangle, \tag{3.24}$$

the expectation values of $\hat{\mathbf{S}}_L^2$ and $\hat{\mathbf{V}}_L^2$ share no such trivial relationship. However their expectation values must be equal in the limit of large M and L since $\hat{\mathbf{S}}_L^2$ is an unbiased estimator of $\hat{\mathbf{V}}_L^2$.

In the remainder of this work, we will restrict our discussion to the expectation values and variances of *coordinate* and *local* operators. We have taken the stationary distribution to be nodeless and positive everywhere. Derivative operators effectively transform the stationary distribution into a new distribution with nodes and negative regions. Such a function cannot be simply interpreted as a standard probability distribution and in this form is equivalent to the fermion problem in quantum Monte Carlo. Since this work focuses on bosonic systems, we leave this matter to be considered in future work.

3.6 The coefficient distribution, characteristic functions, and convolution

Let us now examine in detail the implications for the case where $\hat{A}(\mathbf{x})$ is the identity operator. Consider equation 3.1, letting the expansion for $\tilde{\pi}_N(\mathbf{x})$ contain only a single delta function. This is a one point approximation to $\pi(\mathbf{x})$, where one value of $\mathbf{x} = \mathbf{x}_1$ is randomly drawn from $\pi(\mathbf{x})$. The overlap coefficient is given by

$$\tilde{c}_{0,1} = \int P_0(\mathbf{x})\delta(\mathbf{x} - \mathbf{x}_1)d\mathbf{x} = P_0(\mathbf{x}_1). \tag{3.25}$$

If we repeatedly sample $\pi(\mathbf{x})$ with random single point estimates (snapshots), we build up a set of coefficients belonging to some distribution of coefficients, $C_{0,1}(c)$. In this way, $P_0(\mathbf{x})$ maps the function $\pi(\mathbf{x})$ onto $C_{0,1}(c)$. This provides a way to determine the coefficient distribution *a priori* rather than through direct simulation.

It is often impractical to compute properties directly over the coefficient distribution. Sometimes, $C_{0,1}(c)$ is difficult to obtain, particularly in higher dimensions, since the inverse of $P_0(\mathbf{x})$ is multivalued. If we are only interested in the moments (mean, variance, etc) of the distribution $C_{0,1}(c)$, and not in determining the distribution $C_{0,1}(c)$ itself, we can simplify the analysis considerably. The statistical properties of any operator distribution can be directly computed over the stationary distribution by using $P_0(\mathbf{x})\hat{A}(\mathbf{x})$ to transform integrals over the coefficient distribution into integrals over the stationary distribution. Consider the expected value of c over the $C_{0,1}(c)$ coefficient distribution

$$\langle c \rangle = \int c\, C_{0,1}(c)\, dc. \tag{3.26}$$

Replacing c by $P_0(\mathbf{x})\hat{\mathbf{1}}$ and making the change of variables $C_{0,1}(c)dc = \pi(\mathbf{x})d\mathbf{x}$, we obtain

$$\langle c \rangle = \int P_0(\mathbf{x})\pi(\mathbf{x})\, d\mathbf{x}. \tag{3.27}$$

Likewise, any other operator written in terms of c can be transformed and evaluated over the stationary distribution. For example, the variance of the coefficients $(c - \langle c \rangle)^2$ would be given by

$$\langle c^2 \rangle - \langle c \rangle^2 = \int \left(P_0(\mathbf{x}) - \langle P_0(\mathbf{x}) \rangle \right)^2 \pi(\mathbf{x}) \, d\mathbf{x} = \langle V_{N,M}^2(\hat{1}) \rangle. \tag{3.28}$$

which is just the expectation value of the $\hat{V}_{N,M}^2$ operator.

We now turn our attention to the case where $\pi(\mathbf{x})$ is sampled not with a single point, but with some finite set of points. This discrete version of $\pi(\mathbf{x})$ is expressed in equation 3.1 as a sum of δ-functions. In this case, we obtain coefficients according to

$$\tilde{c}_{0,N} = \int P_0(\mathbf{x}) \frac{1}{N} \sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}_i) \, d\mathbf{x} = \frac{1}{N} \sum_{i=1}^{N} P_0(\mathbf{x}_i) \tag{3.29}$$

where $\tilde{c}_{0,N}$ now comes from the distribution of $C_{0,N}(c)$, the distribution of the mean of N $\tilde{c}_{0,1}$ values. Observations from $C_{0,N}(c)$ are averages of N values of $\tilde{c}_{0,1}$ taken at specific snapshots. As we increase the size of the population, the central limit theorem predicts that the coefficients $\tilde{c}_{0,N}$ will be distributed normally in the limit of large N, with a variance proportional to N^{-1}. For small populations, we *cannot* assume that the coefficients are normally or even symmetrically distributed.

To examine the small N behavior of the coefficient distribution, we invoke the properties of characteristic functions. Characteristic functions are generating functions for probability distributions. They are given by the expected value of $e^{it\mathcal{X}}$, where $\mathcal{X}$ is some random variable taken from the distribution of interest—in this case, the coefficient distribution. Let $\phi_{C_{0,1}}$ denote the characteristic function of the coefficient distribution $C_{0,1}(c)$. We are interested in finding the characteristic function for the average of several random variables all taken from $C_{0,1}(c)$. Thus, we seek the expected value of the quantity

$$\exp\left(\frac{it}{N}\mathcal{C}_{0,N}\right) \tag{3.30}$$

where $\mathcal{C}_{0,N}$ is a sum of single random variables taken from the distribution $C_{0,1}(c)$. However, the characteristic function for a sum of random variables is simply the product of individual characteristic functions

$$\phi_{C_{0,N}} = \prod_{\ell=1}^{N} \phi_{C_{0,1}}. \tag{3.31}$$

Since all of our random variables come from the same coefficient distribution $C_{0,1}(c)$, this product reduces to

$$\phi_{C_{0,N}} = \left(\phi_{C_{0,1}}\right)^N, \tag{3.32}$$

the Nth power of a single characteristic function. Thus, if we know $\phi_{C_{0,1}}$, we can obtain the characteristic function for the distribution $C_{0,N}(c)$. This is accomplished by using the expectation value of the operator defined in equation 3.30, which is simply an inverse Fourier transform. So, to find the distribution of $C_{0,N}(c)$ given $C_{0,1}(c)$, we first take the inverse transform of $C_{0,1}(c)$, raise the result to the desired power to obtain $\phi_{C_{0,N}}$ (equation 3.32), and then use the forward transform.

This process is identical to convolving the distribution with itself N times:

$$\mathcal{C}_{0,N} = \{\mathcal{C}_{0,1} * \mathcal{C}_{0,1} * \ldots * \mathcal{C}_{0,1}\}. \tag{3.33}$$

If analytical formulas for these transforms are available, $C_{0,N}(c)$ may be obtained in closed form, otherwise numerical methods such as the Fast Fourier Transform (FFT) must be employed. Numerical methods have the drawback that large powers of the characteristic function tend to amplify numerical noise. In practice, this method gives reliable distribution functions only for $N < 20$.

The primary reason for pursuing explicit $C_{0,N}(c)$ distributions is to determine, at least qualitatively, how the non-Gaussian character of these operator distributions changes according to dimensionality. For one dimensional distributions, relatively small populations are necessary to achieve central limit theorem behavior. For high dimensional systems where QMC techniques are commonly used, the degree of non-normal character in these distributions is unknown. Furthermore, the population sizes necessary in multiple dimensions to guarantee central limit theorem behavior are also unknown. Thus, a primary focus of this work is to determine the dimensional scaling of operator distributions and accordingly the amount of sampling required to approach the central limit theorem regime.

3.7 Coefficient variance distributions

Just as $C_{0,N}(c)$ describes the distribution of coefficients about their mean value, the variance distribution describes the clustering of variance estimates around the mean value of the variance. For every operator, there are two such distributions: the population variance distribution which is fully determined by moments of π and the sample variance distribution determined by the sampling moments of $\tilde{\pi}$. The population variance distribution has mean $\langle V^2_{N,M}\rangle$ and variance $\langle \mathbf{V}^2_L\rangle$ while the sampling distribution has mean $\langle \bar{S}^2_{N,M}\rangle$ and variance $\langle \bar{\mathbf{S}}^2_L\rangle$. Probing the nature of these distributions is insightful and can indicate when the central limit theorem is applicable and at what point symmetric confidence intervals derived from the normal distribution may be applied. Additionally, this information can aid in the development of new, unbiased and lower variance estimators.

To obtain the population variance distribution of the coefficient operator, we can make use of the convenient change of variables

$$v = (c - \langle c\rangle)^2 \tag{3.34}$$

to transform a coefficient distribution $C_{0,N}(c)$ into the coefficient population variance distribution $V_{0,N}(v)$. Solving this expression for c and applying the change of variables yields

$$V_{0,N}(v) = C_{0,N}(\sqrt{v} + \langle c\rangle)\left|\frac{1}{2\sqrt{v}}\right| + C_{0,N}(-\sqrt{v} - \langle c\rangle)\left|\frac{-1}{2\sqrt{v}}\right|. \tag{3.35}$$

Here, the inverse is multivalued so the use of both the positive and negative roots is required. This technique is general and may be used to derive the variance distribution for any operator, given

the corresponding operator's statistical distribution. The population variance corresponds to the situation where certain parameters, such as the mean and variance of the coefficient distribution, are known exactly. If these parameters must be estimated from the population (as in the case of the sample variance) we lose degrees of freedom. For small populations, this can have a non-trivial effect on the shape, mean, and variance of the sample variance distribution.

Because the sample variance operator, $\hat{S}^2_{M,N}$, involves multiple samples from the coefficient distribution, we cannot obtain its distribution with a simple change of variables. Moreover, the inverse of $\hat{S}^2_{N,M}$ is multivalued, further frustrating a general analytical approach to obtaining this distribution. Consequently, we revert to pure numerical or Monte Carlo sampling to generate sample variance distributions. However, the mean and variance of this distribution can still be obtained analytically from $\langle \bar{S}^2_{N,M} \rangle$ and $\langle \bar{\mathbf{S}}^2_L \rangle$.

3.8 Dimensionality, expectation values and variances

In the previous section, we have considered only properties of the identity operator in one spatial dimension. Let us now consider the properties of an arbitrary k-dimensional coordinate operator expressed as a sum over one dimensional operators

$$\hat{A} = \frac{1}{k} \sum_{i=1}^{k} \hat{A}_i. \tag{3.36}$$

Common positional operators, such as $\hat{A} = \mathbf{x}^2$, can easily be expressed in this form. Since we have chosen to work with multidimensional stationary distributions that are isotropic and uncoupled (equation 3.2), the statistical mean of observed values of $\hat{A}$ in k dimensions is

$$\langle \Omega \rangle_k = \mathcal{I}_{1,1} \mathcal{I}_{1,0}^{k-1} \tag{3.37}$$

where $\hat{\Omega} = P_n \hat{A}$ and $\mathcal{I}_{a,b} = \langle P_n^a \hat{A}^b | \pi \rangle$ are one dimensional integrals in powers of P_n and $\hat{A}$ (proof A.1). Likewise, as demonstrated in proofs A.2 and A.3, the expected value of the variance can be expressed as

$$\langle V^2_{N,M} \rangle_k = \frac{M-1}{M} \langle \bar{S}^2_{N,M} \rangle_k = \frac{1}{N} \left\{ \frac{1}{k} \mathcal{I}_{2,2} \mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k} \mathcal{I}_{2,1}^2 \mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2 \mathcal{I}_{1,0}^{2k-2} \right\}. \tag{3.38}$$

The factor of N is the population size and reflects a random sample taken from that population. Larger populations reduce the variance accordingly. The variance of the population variance is expressed as (proof A.4)

$$\langle \mathbf{V}^2_L \rangle_k = \frac{1}{MN^3} \left[\mu_4 + (2N-3)\mu_2^2 \right], \tag{3.39}$$

where the central moments (μ_b) are defined in terms of the raw moments of $\hat{\Omega}$ as

$$\mu_2 = \langle \Omega^2 \rangle_k - \langle \Omega \rangle_k^2 \tag{3.40}$$

$$\mu_4 = \langle\Omega^4\rangle_k - 4\langle\Omega^3\rangle_k\langle\Omega\rangle_k + 6\langle\Omega^2\rangle_k\langle\Omega\rangle_k^2 - 3\langle\Omega\rangle_k^4 \tag{3.41}$$

and

$$\langle\Omega^2\rangle_k = \frac{1}{k}\mathcal{I}_{2,2}\mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k}\mathcal{I}_{2,1}^2\mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2\mathcal{I}_{1,0}^{2k-2}, \tag{3.42}$$

$$\langle\Omega^3\rangle_k = \frac{1}{k^2}\mathcal{I}_{3,3}\mathcal{I}_{3,0}^{k-1} + \frac{k-1}{k^2}\mathcal{I}_{3,2}\mathcal{I}_{3,1}\mathcal{I}_{3,0}^{k-2} + \frac{k-1}{k}\mathcal{I}_{3,1}^3\mathcal{I}_{3,0}^{k-3}, \tag{3.43}$$

$$\langle\Omega^4\rangle_k = \frac{1}{k^3}\mathcal{I}_{4,4}\mathcal{I}_{4,0}^{k-1} + \frac{k-1}{k^3}\left(\mathcal{I}_{4,3}\mathcal{I}_{4,1} + \mathcal{I}_{4,2}^2\right)\mathcal{I}_{4,0}^{k-2} + \frac{k-1}{k^2}\mathcal{I}_{4,2}\mathcal{I}_{4,1}^2\mathcal{I}_{4,0}^{k-3} + \frac{k-1}{k}\mathcal{I}_{4,1}^4\mathcal{I}_{4,0}^{k-4}. \tag{3.44}$$

The factor of M in equation 3.39 is related to the random sample size, or the number of snapshots of the population that we may average over. Including more snapshots reduces the statistical uncertainty in estimates of the variance. Similarly, the sampling variance of the sample variance in k dimensions is (proof A.5)

$$\langle\bar{\mathbf{S}}_L^2\rangle_k = \left(\frac{L}{L-1}\right)\frac{1}{MN^3}\left[\mu_4 + \left(\frac{2NM}{M-1} - 3\right)\mu_2^2\right], \tag{3.45}$$

and clearly tends to the same limit as $\langle\mathbf{V}_L^2\rangle_k$ as M and L become large.

As we shall see, these formulas, all with exponential dependence on k, have profound implications for the performance of QMC methods in medium and high dimensional systems. As we will discuss in the next section, the use of importance sampling can help to dramatically suppress (but not entirely eliminate) this exponential dependence.

3.9 Extension to importance sampling

We can introduce an importance sampling function, $\psi_T(\mathbf{x})$, such that the transformation

$$\hat{\mathcal{A}}' = \psi_T^{-1}\hat{\mathcal{A}}\psi_T \tag{3.46}$$

leads to a new operator $\hat{\mathcal{A}}'$ having an expectation value and statistical variance identical to the untransformed operator, $\hat{\mathcal{A}}$. Taking equation 3.7, and inserting our transformed operator, we obtain

$$\int \frac{P_0}{\psi_T}\hat{\mathcal{A}}\psi_T(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}. \tag{3.47}$$

In importance sampling, we lump both $\psi_T(\mathbf{x})$ and $\pi(\mathbf{x})$ into a new stationary distribution $\pi'(\mathbf{x})$ and redefine our projector to be

$$P_0' = \frac{P_0}{\psi_T}, \tag{3.48}$$

Clearly, the expectation value of $\hat{\mathcal{A}}$ (computed from equation 3.8)

$$\langle\hat{\mathcal{A}}\rangle = \langle P_0'\hat{\mathcal{A}}|\pi'(\mathbf{x})\rangle = \langle P_0\hat{\mathcal{A}}|\pi(\mathbf{x})\rangle \tag{3.49}$$

remains unchanged by this redefinition, but the expected value of the variance becomes

$$\langle V_{N,M}^2 \rangle = \langle (P_0'\hat{A})^2 | \pi'(\mathbf{x}) \rangle - \langle P_0'\hat{A} | \pi'(\mathbf{x}) \rangle^2 \neq \langle (P_0\hat{A})^2 | \pi(\mathbf{x}) \rangle - \langle P_0\hat{A} | \pi(\mathbf{x}) \rangle^2. \tag{3.50}$$

As $\psi_T(\mathbf{x}) \to P_0(\mathbf{x})$, P_0' approaches unity and the statistical variance is transformed into the quantum mechanical variance, $\langle \Omega(\hat{A}^2 - \langle A \rangle^2) \rangle$. One consequence is that zero-variance operators (such as the energy) will also have no statistical error in the limit of perfect importance sampling. This result is true regardless of the dimensionality of the system and is good news for the computation of certain observables. However, for other arbitrary operators that do not possess a zero-variance principle, importance sampling can actually *increase* the statistical error of raw estimators.

Assume now that ψ_T depends on some tunable parameter α, such that when $\alpha = 0$, $P_0' = P_0$ and when $\alpha = 1$, $P_0' = 1$. We will denote this tunable trial function ψ_T^α and build this parameter dependence into a new, more general integration scheme

$$\langle \Omega \rangle^\alpha = \int \hat{\Omega}_n^\alpha(\mathbf{x}) \pi'(\mathbf{x}) d\mathbf{x}, \tag{3.51}$$

where equation 3.8 is recovered in the limit $\alpha = 0$. Likewise, for a particular value of α, there will be a unique set of projectors $\langle \hat{\Omega}_n^\alpha |$.

Importance sampling has the overall effect of reducing the statistical variance of an operator by altering the stationary distribution and projectors. Quantum mechanical observables remain unaffected by importance sampling. Coefficient and operator distributions tend to get narrower and more sharply peaked around their true quantum mechanical values as $\alpha \to 1$. However, it is unknown whether these importance sampled distributions remain asymmetric and skewed as $\alpha \to 1$, or whether they evolve toward Gaussian distributions. QMC simulations that rely primarily on importance sampling to reduce the statistical variance could significantly underestimate the uncertainty in a calculation if these distributions are non-Gaussian.

3.10 Dimensionality, importance sampling, and pseudo zero-variance principles

Earlier, we demonstrated that for operators of the form

$$\hat{A} = \frac{1}{k} \sum_{i=1}^{k} \hat{A}_i \tag{3.52}$$

the k-dimensional expectation value could be expressed as

$$\langle \Omega(\hat{A}) \rangle_k = \mathcal{I}_{1,1} \mathcal{I}_{1,0}^{k-1}. \tag{3.53}$$

For properly normalized projectors, $\mathcal{I}_{1,0} = 1$ and $\langle \Omega \rangle_k$ is constant and independent of dimension. Likewise, if we had chosen not to include the factor $1/k$ in our definition of $\hat{\mathcal{A}}$, the expectation value

$$\langle \Omega(k\hat{\mathcal{A}}) \rangle_k = k\mathcal{I}_{1,1}\mathcal{I}_{1,0}^{k-1} \tag{3.54}$$

would increase with dimension, but not exponentially. In both cases, importance sampling does not affect the expectation value of the operator. Now, performing a similar analysis on the k-dimensional variance

$$\langle V_{N,M}^2 \rangle_k = \frac{1}{N}\left\{ \frac{1}{k}\mathcal{I}_{2,2}\mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k}\mathcal{I}_{2,1}^2\mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2\mathcal{I}_{1,0}^{2k-2} \right\} \tag{3.55}$$

and taking the limit of perfect importance sampling ($P_0' = 1$), we obtain

$$\langle V_{N,M}^2(\hat{\mathcal{A}}) \rangle_k = \frac{1}{kN}\left(\langle \Omega^2(\hat{\mathcal{A}}) \rangle^2 - \langle \Omega(\hat{\mathcal{A}}) \rangle^2 \right) \tag{3.56}$$

and

$$\langle V_{N,M}^2(k\hat{\mathcal{A}}) \rangle_k = \frac{k}{N}\left(\langle \Omega^2(\hat{\mathcal{A}}) \rangle^2 - \langle \Omega(\hat{\mathcal{A}}) \rangle^2 \right). \tag{3.57}$$

There are several important implications here. First, using our rather contrived choice for $\hat{\mathcal{A}}$, we see that the k-dimensional variance decreases as k^{-1}. Thus, in the case where the expectation value of $\hat{\mathcal{A}}$ is engineered to be constant or decreasing with dimension, we can expect that the variance will act similarly. This leads to a pseudo zero-variance principle which indicates that for near-perfect importance sampling, the variance will decrease as the dimensionality increases. Consequently, for cleverly chosen operators in high dimensional systems, this may provide a quick trick for variance control without the expense of explicitly designing zero-variance operators. Second, for more realistic operators whose expectation values increase with dimension, it is clear that even for perfect importance sampling, the variance will also increase with dimension.

One might be tempted to think that in taking $\mathcal{I}_{1,0}$ to be equal to unity for a properly normalized projector we have defeated the exponential dependence of the variance. However, this is not the case. The one dimensional integral $\mathcal{I}_{2,0}$ (not $\mathcal{I}_{1,0}$) is the actual source of the exponential dependence on dimension. Thus, the one dimensional functional form of $\Omega(\hat{\mathcal{A}})$ determines the degree of exponential dependence. Only in the limit of perfect importance sampling where the projector P_0' is constant and equal to unity can the dimensional dependence be fully eliminated.

3.11 Summary

In this chapter, we have developed a general theoretical formalism for characterizing the statistical error of arbitrary observables due to a finite sampling population. Formulas for the population and sampling variances as well as for the second moments of these variances have been derived. We have shown that characteristic functions provide a method for explicitly examining the evolution of statistical error far from the central limit theorem regime and are useful for better understanding the practical population size requirements required in QMC investigations. Most importantly, we

have considered the effects of dimensionality and importance sampling on the nature of these operator distributions and find an inescapable exponential scaling of the error with dimension[59–62] for all but perfect importance sampling. This puts strenuous demands on the quality of trial functions required for accurate calculations in systems of large dimensionality.

For importance sampling, we find that as the trial function Ψ_T approaches the true wavefunction, the statistical error becomes equal to the expectation value of the quantum mechanical variance. For zero-variance operators, this means that the statistical error will go to zero in the limit of perfect importance sampling. For arbitrary operators, better importance sampling does not guarantee a decrease in the statistical error. We also demonstrate that for certain classes of operators, there exists a pseudo-zero-variance principle in high dimensional systems that can guarantee vanishing statistical variance in the limit of perfect importance sampling.

With suitable modification, the formulas and underlying theory presented in this chapter may be used to assess *a priori* the expected statistical error in a calculation as well as recommend suitable population sizes and levels of importance sampling necessary to bound the error at a given level. The predictive power of this model and its practical implementation is a subject of ongoing research.

Chapter 4

Application to the k-dimensional Harmonic Oscillator

4.1 Introduction

In this chapter we apply the theoretical ideas developed in the previous chapter to the k-dimensional simple harmonic oscillator (SHO). As a first approximation to an interatomic potential energy curve, the k-dimensional SHO model is expected to share scaling properties representative of atomic clusters. We examine the effects of ground state properties on importance sampling, population size, and dimensionality for several operators including two important zero-variance operators, the local energy operator E_L and the variance of the local energy, $(\hat{E}_L - \langle E_L \rangle)^2$.

4.2 The stationary distribution and projection operators

For the one dimensional harmonic oscillator with unit force constant, we choose the stationary distribution to be proportional to the ground state eigenfunction

$$\pi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{4.1}$$

where $\pi(x)$ is normalized to integrate to unity. For importance sampling, we choose the trial wavefunction

$$\psi_T^{\alpha}(x) = \frac{\sqrt{\alpha+1}}{\sqrt{2}\pi^{1/4}} e^{-\alpha x^2/2} \tag{4.2}$$

so that our new stationary distribution becomes

$$\pi(x) = \sqrt{\frac{\alpha+1}{2\pi}} e^{-(\alpha+1)x^2/2}. \tag{4.3}$$

We select our projection operators to be suitably renormalized harmonic oscillator wavefunctions divided by the trial wavefunction

$$P_v(x) = \mathcal{N}_v \frac{\psi_v(x)}{\psi_T^\alpha(x)} = \mathcal{N}_v \frac{\sqrt{2}}{\sqrt{\alpha+1}} \widetilde{\mathcal{H}}_v(x) e^{-(1-\alpha)x^2/2}, \tag{4.4}$$

where $\widetilde{\mathcal{H}}_v(x)$ is the nth Hermite polynomial in x, $\mathcal{H}_v(x)$, multiplied by the normalization factor $(2^v v!)^{-1/2}$. From equation 3.4, we have $\mathcal{N}_0 = 1$. For $v > 0$, we cannot determine $\mathcal{N}_v$ in an analogous way, since these integrals are all zero by orthogonality. We must instead choose normalizations based on the matrix elements of operators other than $\hat{1}$. In this study, normalizations for all states are taken to be identical such that $\mathcal{N}_0 = \mathcal{N}_v = 1$.

4.3 The coefficient distribution

Taking the ground state projector P_0, normalized as above, we determine the ground state ($v = 0$) coefficient distribution for a population size $N = 1$ as a function of dimension k and importance sampling parameter α to be (proof A.6)

$$C^{\alpha,k}_{v=0,N=1}(c) = \frac{(\alpha+1)^k w^{\alpha/(\alpha-1)}}{2^{k-1}(\alpha-1)\Gamma(k/2)} \left(\frac{\ln w}{1-\alpha}\right)^{(k-2)/2} \tag{4.5}$$

where

$$w = c^2 \left(\frac{\alpha+1}{2}\right)^k. \tag{4.6}$$

The coefficient distribution for dimensions 1, 2, 3 and 7 with $\alpha = 0$ are shown in figure 4.1. As dimension increases, the trend is for the coefficient distribution to be become more sharply peaked at $c = 0$ and more skewed with a long tail reaching out to large values of c. Curiously, the one and two dimensional distributions are the most unusual, the former because of the singularity at $c = \sqrt{2}$ and the latter because it is exactly uniform over the support $0 \leq c < 2$. By setting $\alpha = 0$ in equation 4.5, we see that the support of these distributions is expanding as $2^{k/2}$. For all values of α, the mean remains equal to unity; however, in higher dimensions the increased spatial extent of the distributions forces the variance to be larger. This directly correlates with inadequate Monte Carlo sampling in high dimensional space. Very few points located in the skewed tail tend to contribute appreciably to the average. Accordingly, efficiency is reduced since most sampling points provide insignificant contributions to the average.

The shape of these distributions is profoundly affected by the population size and the importance sampling parameter α. Figures 4.2, 4.3, 4.4 and 4.5 show the effects of importance sampling and population size on distributions of dimensions 1, 2, 3 and 7, respectively. In these figures, panel (a) shows what happens to the ground state coefficient distribution as we gradually turn on importance sampling, with $\alpha \to 1$ in increments of 0.25. We see that the effect of importance sampling on these distributions is to make them more peaked around the expected coefficient value

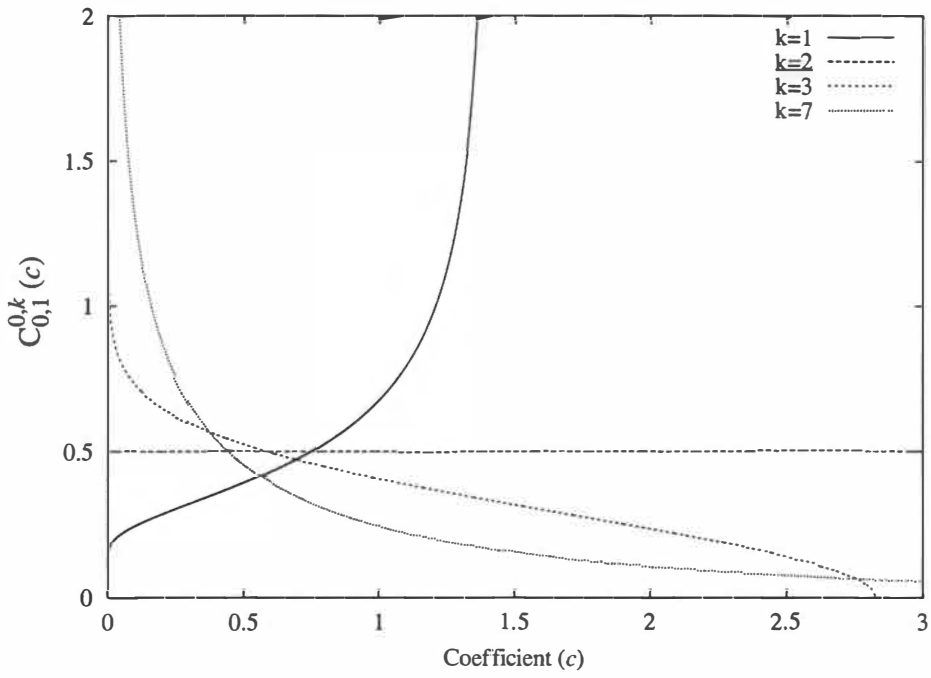


Figure 4.1: Ground state coefficient distributions for 1,2,3 and 7 dimensions, no importance sampling and a population size of 1.

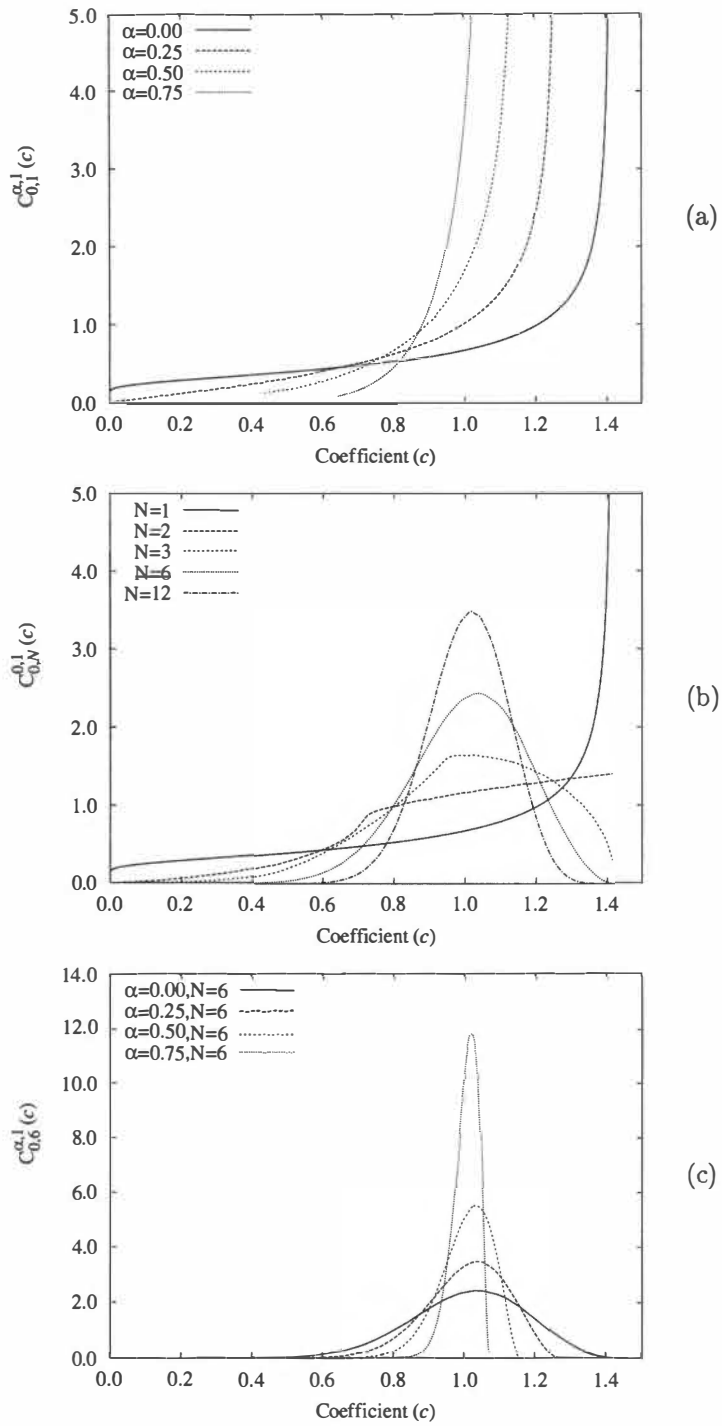


Figure 4.2: One dimensional ground state coefficient distributions for (a) various levels of importance sampling ($\alpha = 0.0, 0.25, 0.5, 0.75$), (b) several population sizes ($N = 1, 2, 3, 6, 12$), and (c) importance sampling with the population $N = 6$.

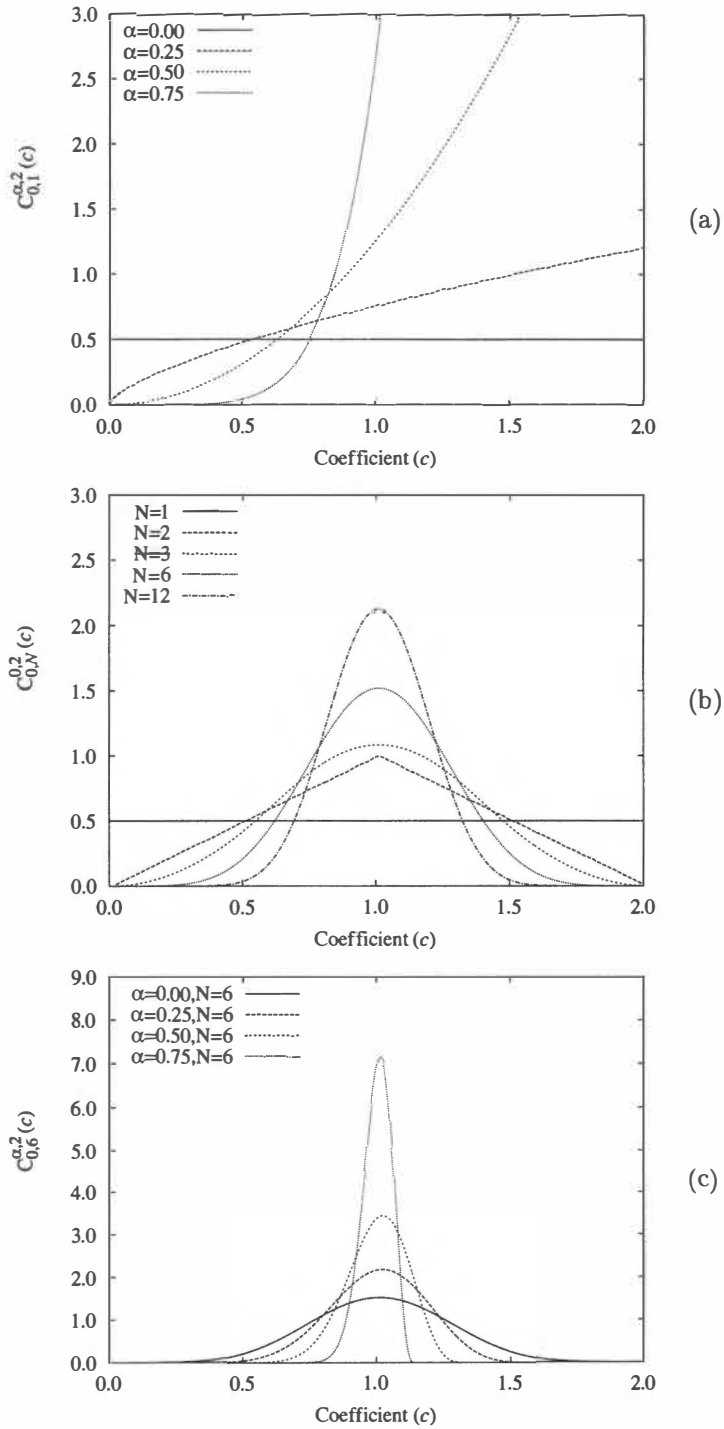


Figure 4.3: Two dimensional ground state coefficient distributions for (a) various levels of importance sampling ($\alpha = 0.0, 0.25, 0.5, 0.75$), (b) several population sizes ($N = 1, 2, 3, 6, 12$), and (c) importance sampling with the population $N = 6$.

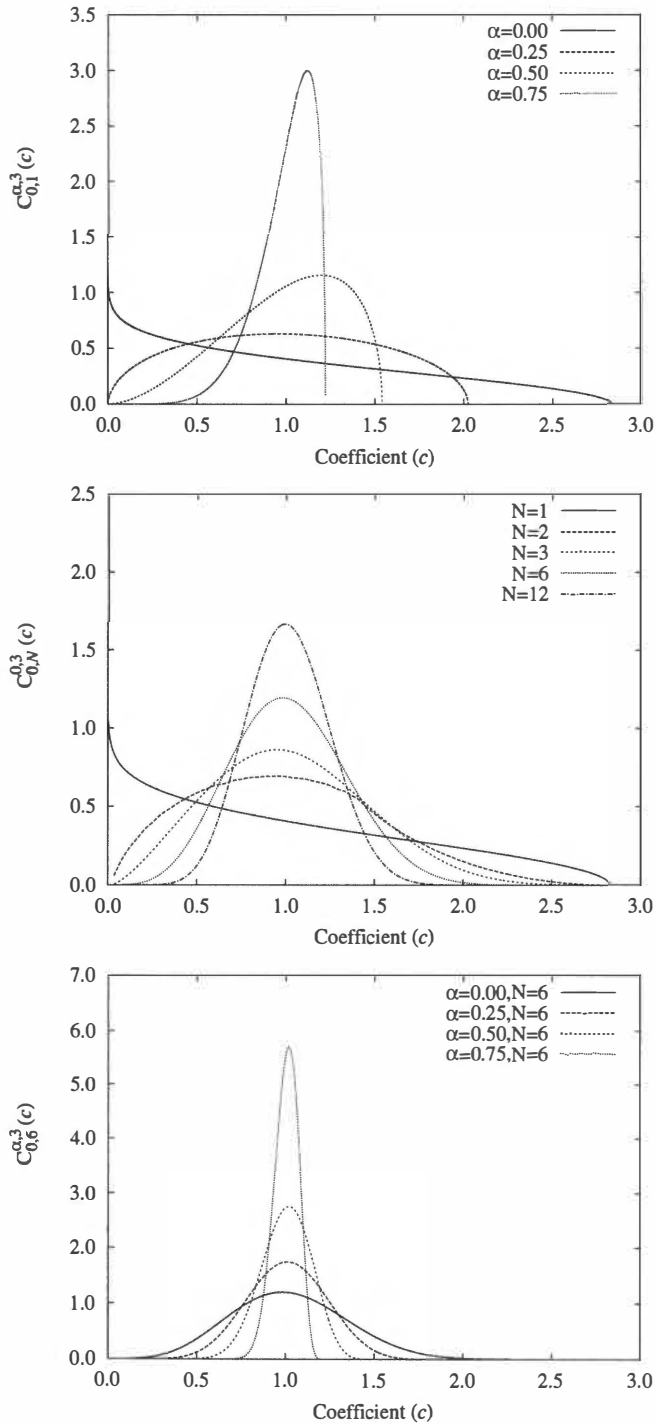


Figure 4.4: Three dimensional ground state coefficient distributions for (a) various levels of importance sampling ($\alpha = 0.0, 0.25, 0.5, 0.75$), (b) several population sizes ($N = 1, 2, 3, 6, 12$), and (c) importance sampling with the population $N = 6$.

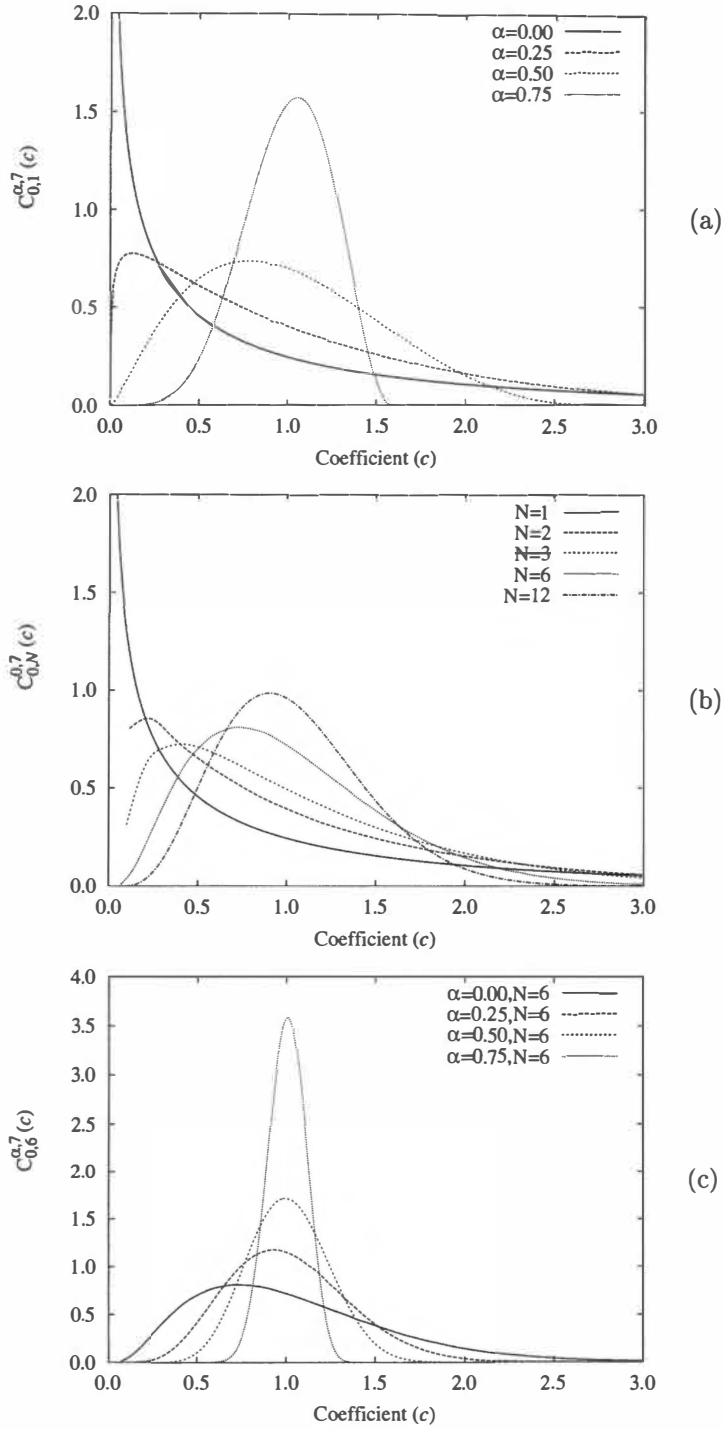


Figure 4.5: Seven dimensional ground state coefficient distributions for (a) various levels of importance sampling ($\alpha = 0.0, 0.25, 0.5, 0.75$), (b) several population sizes ($N = 1, 2, 3, 6, 12$), and (c) importance sampling with the population $N = 6$.

of 1.0. Importance sampling also reduces the extent of the support, so that the tails of the importance sampled distributions are not as long as the tails of non-importance sampled distributions. Strikingly, all distributions approach a delta function asymmetrically, with one side more skewed than the other. Even in the two dimensional case, importance sampling introduces asymmetry into a perfectly symmetric initial distribution. Depending on the extent of this asymmetry, it could be misleading to report a symmetric standard error or confidence interval for the mean of such a distribution.

In panel (b), the coefficient distribution is plotted for various population sizes, where $C_{0,N}^{\alpha,\boldsymbol{k}}(c)$ is obtained from $C_{0,1}^{\alpha,\boldsymbol{k}}(c)$ via equation 3.33 through the numerical application of the discrete Fast Fourier Transform (FFT). With increasing N, the trend is toward a more Gaussian shaped distribution, centered around the mean coefficient value of 1.0. With increasing dimension at a fixed value of N, we see that the distributions become wider and are less Gaussian shaped. Thus identical populations will give larger variances in higher dimensions and the corresponding coefficient distributions will be more skewed. Practically speaking, one must employ a larger population in higher dimensions to maintain the same statistical error. Panel (c) shows a combination of population size and importance sampling effects. As α increases at $N = 6$, we see the distributions quickly sharpen up, with population size fighting importance sampling to keep the distribution symmetric and roughly bell shaped. At the largest values of α, skewness wins out over population averaging.

4.4 The coefficient variance distribution

As described in section 3.7, the population variance distribution may be obtained from the coefficient distribution by a simple change of variables (equation 3.34) while the sample variance distribution for the harmonic oscillator must be determined by Monte Carlo simulation. Figures 4.6, 4.7, 4.8 and 4.9 show the population and sample variance distributions as functions of population size, snapshots, and importance sampling. From these plots, several trends are immediately obvious. As the population size (N) increases, the variance distribution narrows and moves closer to zero (figure 4.6). Larger populations tend toward normal statistics, so we expect the variance to be roughly distributed as χ^2 with M degrees of freedom. This is given directly by applying the change of variables technique to a normally distributed coefficient distribution. Figure 4.7 shows the one dimensional population variance distribution, $V_{0,N,M}^{\alpha,\boldsymbol{k}}(v)$, for various levels of importance sampling. As importance sampling is turned on, the width of the distribution contracts rapidly toward zero. At full importance sampling, the distribution collapses into a delta function at zero. Curiously, the population variance distribution for $N = 1, M = 1$ exhibits a very sharp feature near the mean $(1 - 2/\sqrt{3})$. As the population size is increased (figure 4.8), this feature disappears. Since the sample variance distribution, $S_{0,N,M}^{\alpha,\boldsymbol{k}}(s)$, does not exist for $M = 1$ (we need at least two independent snapshots to compute a sample variance), this feature is unique to the population variance distribution. Increasing the number of snapshots (M) sharpens up the variance distribution around its mean for both the population (figure 4.8) and sample variance (figure 4.9)

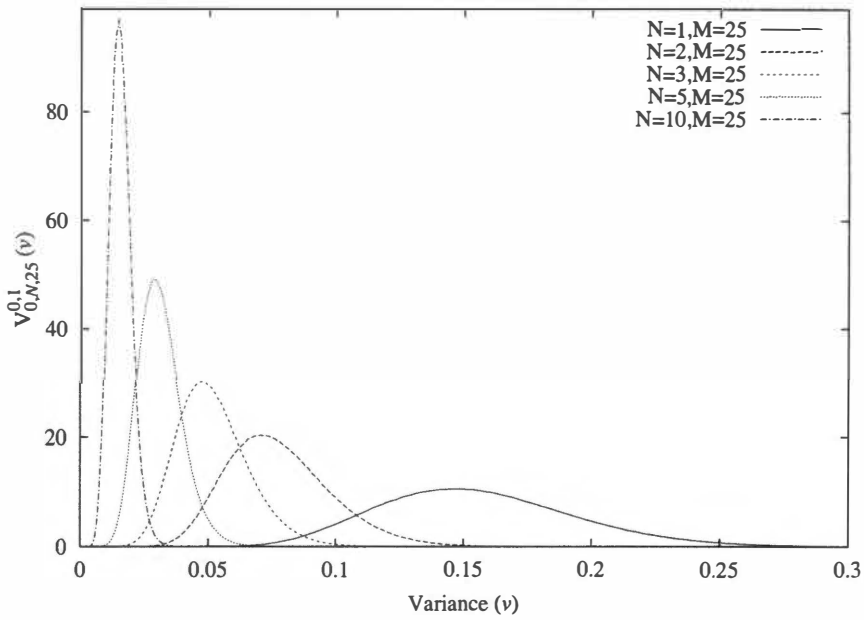


Figure 4.6: Combination effect of population size ($N = 1, 2, 3, 5, 10$) and snapshots ($M = 25$) on the 1D ground state population variance distribution.

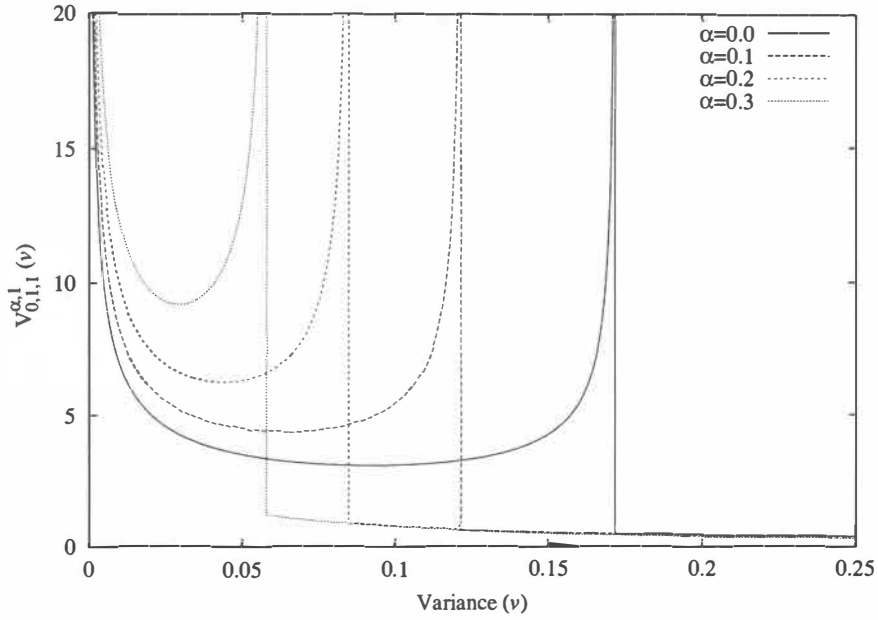


Figure 4.7: Effect of importance sampling ($\alpha = 0.0, 0.1, 0.2, 0.3$) on the 1D ground state population variance distribution for a population size of 1 and $M = 1$.

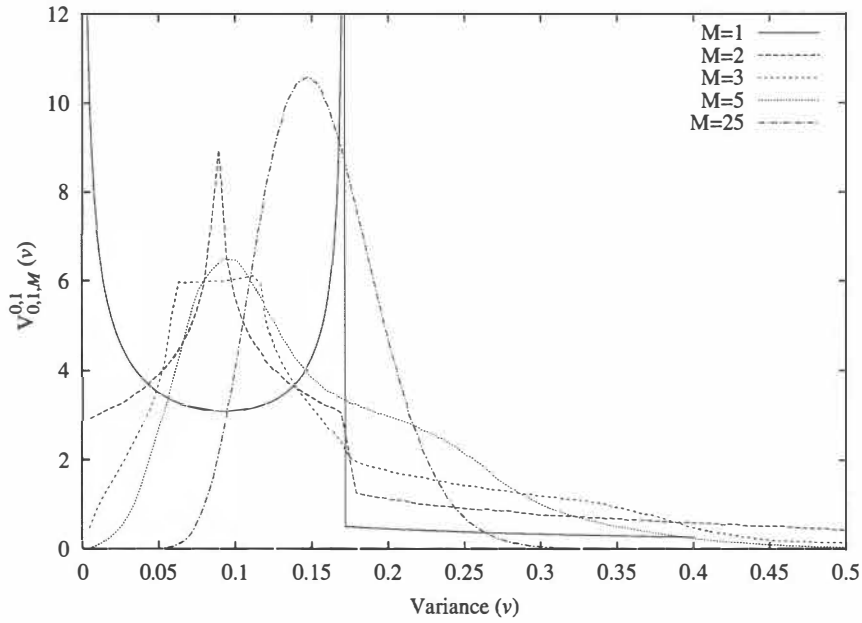


Figure 4.8: Effect of snapshots ($M = 1, 2, 5, 25$) on the 1D ground state population variance distribution for a population size of 1.

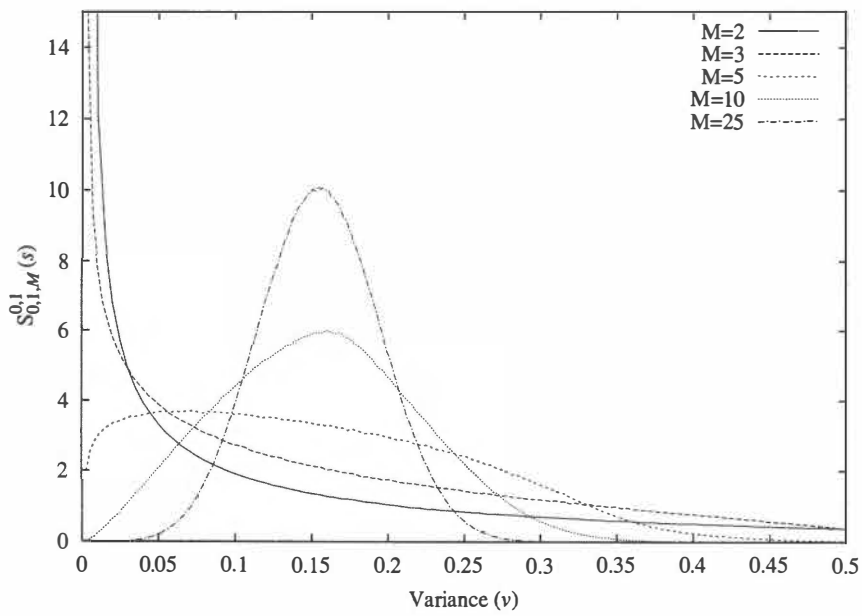


Figure 4.9: Effect of snapshots ($M = 1, 2, 5, 25$) on the 1D ground state sample variance distribution for a population size of 1.

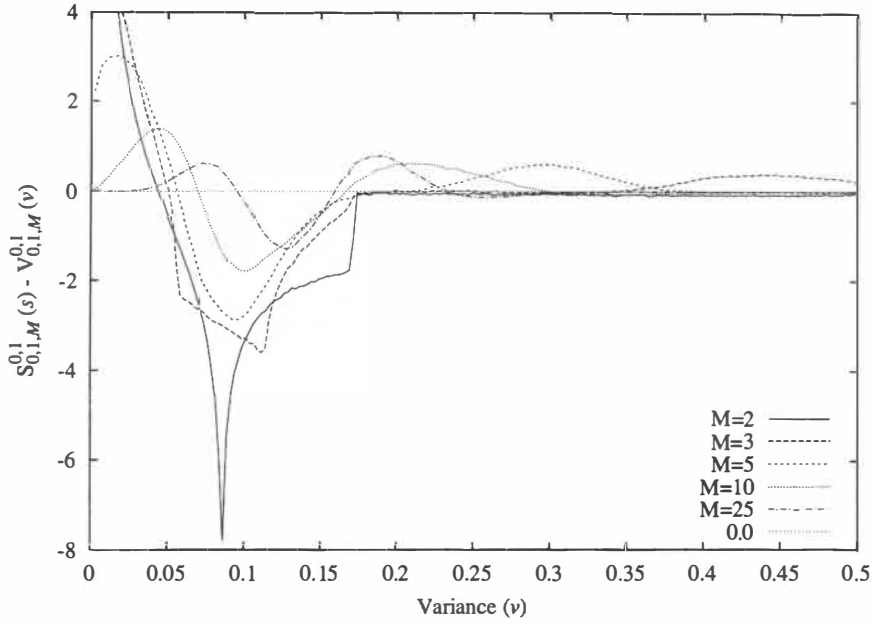


Figure 4.10: Difference between the sample variance distribution and the population variance population for various sampling sizes ($M = 2, 3, 5, 10, 25$) for the 1D ground state.

distributions. The sharpening tends toward a narrow normal distribution as M is increased.

Figure 4.10 shows the difference between the population and sample variance distributions. Positive regions indicate where $S^{0,1}_{0,1,M}(s)$ is greater than $V^{0,1}_{0,1,M}(v)$. For all values of M, the sample variance is more diffuse, deficient near the true mean and increased on either side of the mean, compared to the population variance distribution. Thus, the sample variance distribution has a larger statistical variance than the population variance distribution, in agreement with the expectation values of $\langle \bar{\mathbf{S}}^2_L \rangle$ and $\langle \mathbf{V}^2_L \rangle$.

4.5 Variances of the coefficient operator

Now that we have examined the behavior and shapes of distributions of coefficients, let us now summarize their essential characteristics as functions of importance sampling, dimension, population size, and number of snapshots. These results come from the expectation values $\langle V^2_{N,M} \rangle^\alpha_k$ and $\langle \bar{S}^2_{N,M} \rangle^\alpha_k$ (equation 3.38), $\langle \mathbf{V}^2_L \rangle^\alpha_k$ (equation 3.39), and $\langle \bar{\mathbf{S}}^2_L \rangle^\alpha_k$ (equation 3.45).

Figure 4.11 shows the variance of the coefficient distribution decreasing smoothly for all dimensions as we turn on importance sampling. As $\alpha \to 1$, we see monotonic, exponential decay to zero. The statistical variance increases with increasing dimension. Curves for different values of k never cross except at $\alpha = 1$, where the variance becomes identically zero in all dimensions.

To better illustrate the dimensional dependence of the coefficient operator variance, we have plotted the natural log of $\langle V^2_{N=1,M} \rangle$ vs dimension in figure 4.12. For the first few dimensions

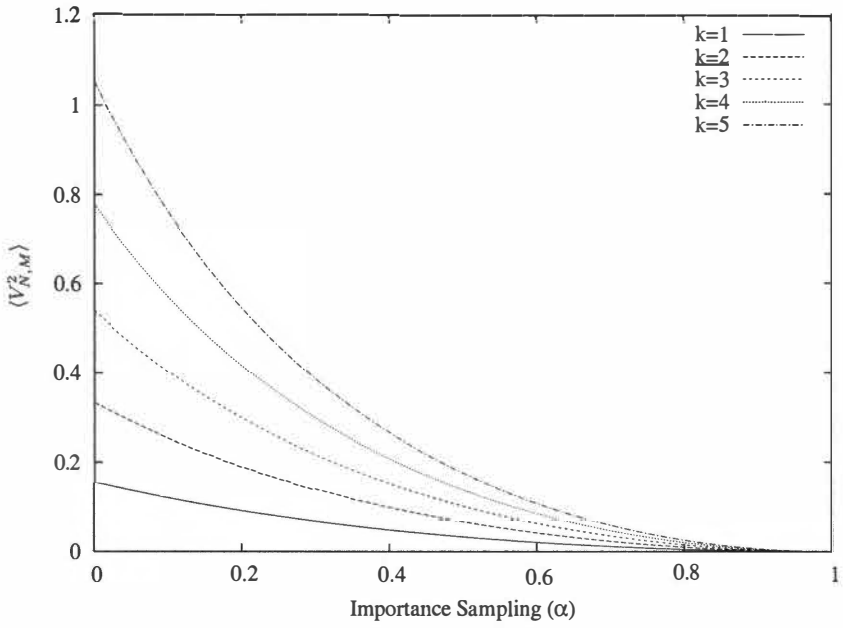


Figure 4.11: Variance of the coefficient operator ($\hat{1}$) vs importance sampling (α) for various dimensions ($k = 1, 2, 3, 4, 5$).

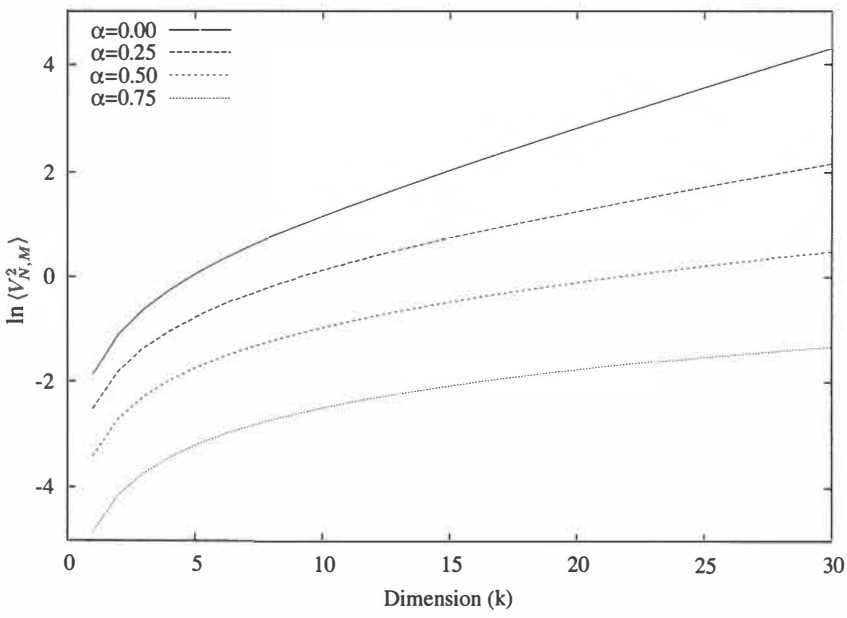


Figure 4.12: Natural logarithm of the variance of the coefficient operator ($\hat{1}$) vs dimension (k) for various levels of importance sampling ($\alpha = 0.0, 0.25, 0.50, 0.75$).

we see a rapid increase of the error. For large values of dimension the behavior is linear and indicates pure exponential growth. This is particularly bad news for the scaling of the error in quantum Monte Carlo methods. While it is true that the rate of convergence in the variance is proportional to N^{-1}, the actual error increases exponentially with dimension. Either strong importance sampling is required, or much larger population sizes must be employed to combat this exponentially increasing error.

Additional curves for various levels of importance sampling are plotted in figure 4.12 and indicate that importance sampling dramatically improves the scaling of the error with dimension. For thirty dimensions with $\alpha = 0.75$ importance sampling, the error is nearly the same as for one dimension and no importance sampling. We see both an overall suppression of the error (vertical displacement of the curve) as well as a weaker dependence on dimension (flatter slope) for increased importance sampling.

For population size (not shown), the variance will decrease according to N^{-1} (equation 3.38). On a log plot, this has the effect of vertically displacing the curve by $-\ln N$, but does not alter the strong dependence on dimension. For example, to reduce the variance in thirty dimensions to a level comparable with that achieved with $\alpha = 0.75$ importance sampling only by increasing the population size, it would be necessary to have a population more than 400 times larger. Of course, in actual simulations, we are typically concerned with reducing the sample variance, which is always larger in magnitude than the population variance. However, in some cases such as wavefunction optimization according to the second moment of the local energy, these population variance distributions will come into play. Figure 4.13 shows how much larger the sample variance is in comparison with the population variance (equivalent to $M = \infty$) for various values of M. As dimension increases, the difference between $\langle \bar{S}^2_{N,M} \rangle^\alpha_k$ and $\langle V^2_{N,M} \rangle^\alpha_k$ gets larger. This means that in higher dimensions, more snapshots are required to ensure that $\langle \bar{S}^2_{N,M} \rangle^\alpha_k$ is not too much larger than $\langle V^2_{N,M} \rangle^\alpha_k$

We can also look at general trends for the variance of estimates of the coefficient variance. In figure 4.14 we have plotted $\langle \mathbf{V}^2_L \rangle$ as a function of trial wavefunction quality. The behavior of $\langle \mathbf{V}^2_L \rangle$, is similar to that of $\langle V^2_{N,M} \rangle$. Here, the initial variances are larger in magnitude than those of $\langle V^2_{N,M} \rangle$ but decrease rapidly as $\alpha \to 1$. Thus, for modest importance sampling, the error in the variance will be manageably small. Like $\langle V^2_{N,M} \rangle$, $\langle \mathbf{V}^2_L \rangle$ also tends to zero as $\alpha \to 1$.

In figure 4.15, we see $\ln \langle \mathbf{V}^2_L \rangle$ plotted against dimension. At large dimensions we see exponential growth of the variance, similar to that of $\ln \langle V^2_{N,M} \rangle$. At small dimensions, the behavior is more complex due to the nature of equation 3.39. As before, importance sampling significantly reduces the statistical error for the larger dimensions by displacing the curves and by flattening the slopes. Figure 4.16 shows how the population size affects $\ln \langle \mathbf{V}^2_L \rangle$. Here, there is not simply a vertical shift in the curves since the N dependence of equation 3.39 is not proportional to N^{-1}. Instead, equation 3.39 has terms proportional to both N^{-2} and N^{-3}, yielding a vertical displacement as well as a slight change in the dimensional dependence of the curves as we increase the population size.

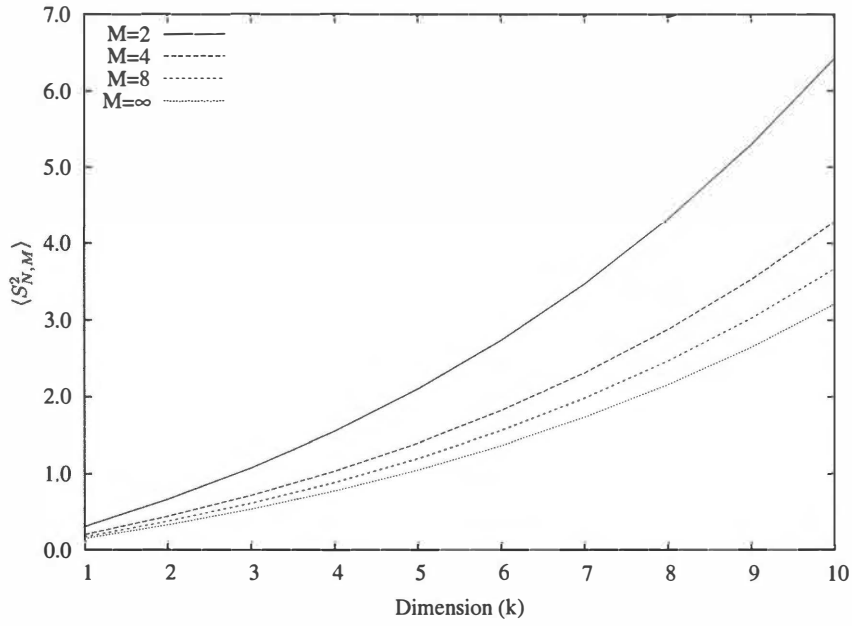


Figure 4.13: Sample variance of the coefficient operator ($\hat{1}$) vs dimension (k) for various numbers of snapshots ($M = 2, 4, 8, \infty$).

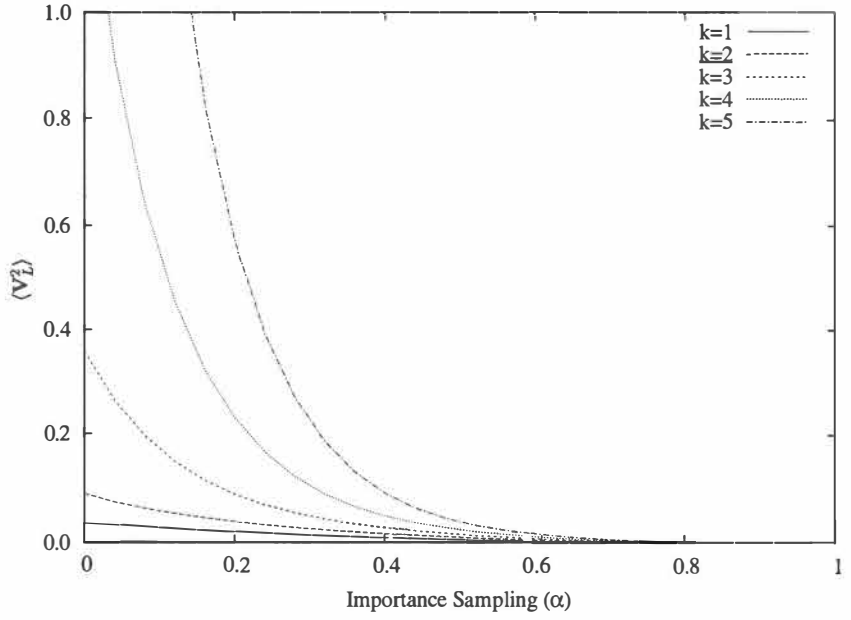


Figure 4.14: Variance of the coefficient variance operator ($\hat{1}$) vs importance sampling (α) for various dimensions ($k = 1, 2, 3, 4, 5$).

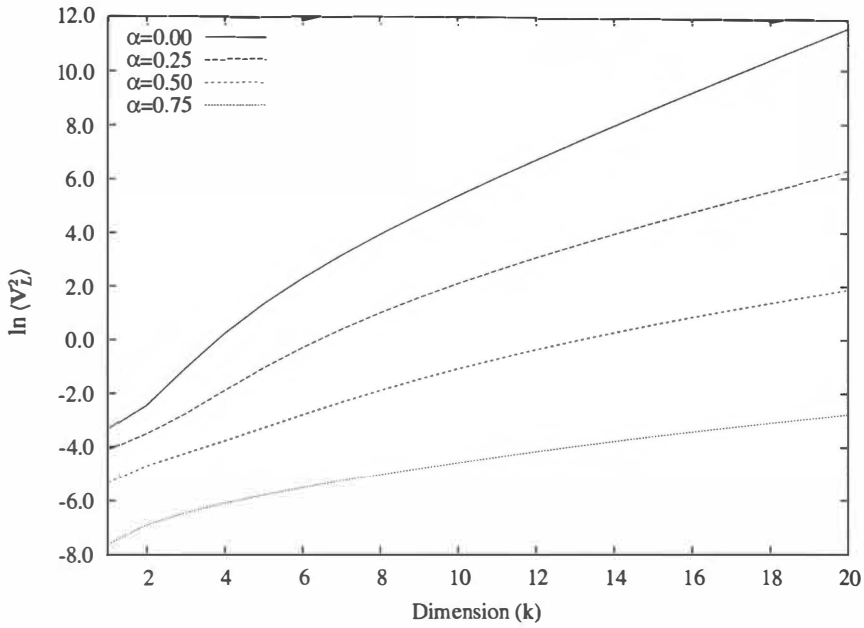


Figure 4.15: Natural logarithm of the variance of the coefficient variance operator ($\hat{1}$) vs dimension (k) for various levels of importance sampling ($\alpha = 0.0, 0.25, 0.50, 0.75$).

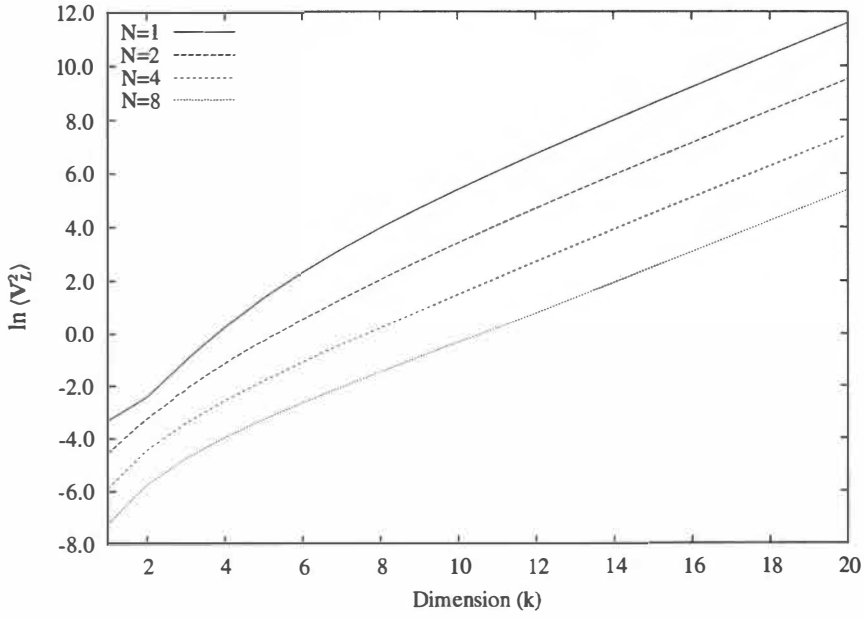


Figure 4.16: Natural logarithm of the variance of the coefficient variance operator ($\hat{1}$) vs dimension (k) for various population sizes ($N = 1, 2, 4, 8$).

4.6 Variances of the x^2 operator

Since many of the trends observed for the coefficient operator carry over or are similar for the $\mathbf{x}^2$ operator, we will focus only on essential differences. For the coefficient operator, we saw an initial growth of the variance at low dimensionality followed by pure exponential growth at large dimension (figure 4.12). For the $\mathbf{x}^2$ operator (figure 4.17), we again see exponential growth at large dimension. However, at small dimensions, we see an initial *decrease* to some minimum. The position of this minimum shifts to higher dimensions as we increase the degree of importance sampling. The appearance of the minimum is primarily a consequence of the first term of equation 3.38. In the case of the coefficient operator, this first term is fully canceled by part of the second term, and results in no minimum. Here, no such cancellation occurs. As we tune α toward unity, we quench the exponential dependence on k and suppress the overall variance.

The plot of $\langle V^2_{N,M} \rangle$ vs importance sampling (figure 4.18) is equally interesting. For small dimensions, contrary to what one might expect, better importance sampling does not necessarily lead to a smaller statistical variance. Moreover, some finite variance remains even with perfect importance sampling. This is in agreement with the work of Assaraf and Caffarel,[53] in which they assert that arbitrary operators do not possess a zero-variance property. We find this principle to apply in the case of the $\mathbf{x}^2$ operator. However, at large dimensions the minimum variance is achieved at values of α close to unity. This may imply the existence of an *effective* zero-variance property for certain operators in high dimensional systems for which good trial wavefunctions are available.

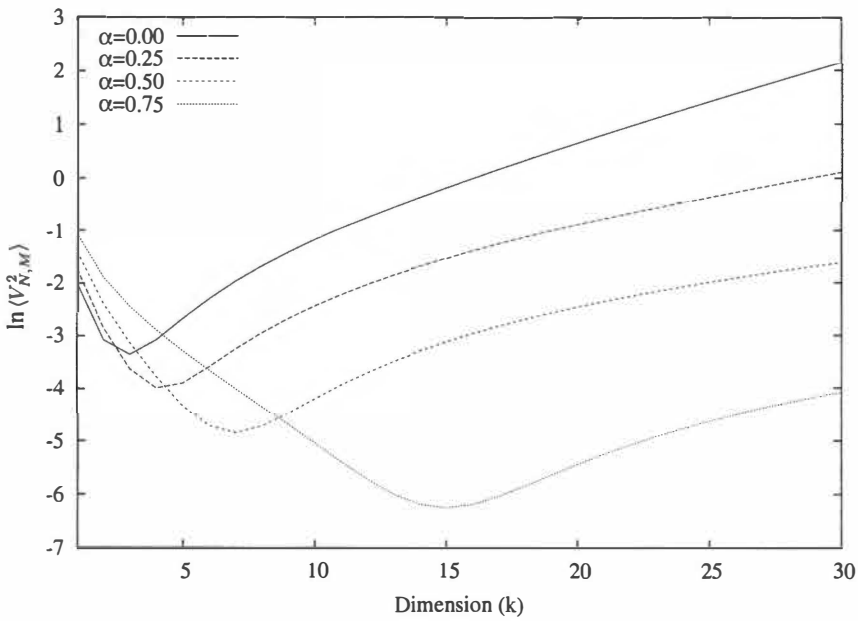


Figure 4.17: Natural logarithm of the variance of the $\mathbf{x}^2$ operator vs dimension (k) for various levels of importance sampling ($\alpha = 0.0, 0.25, 0.50, 0.75$).

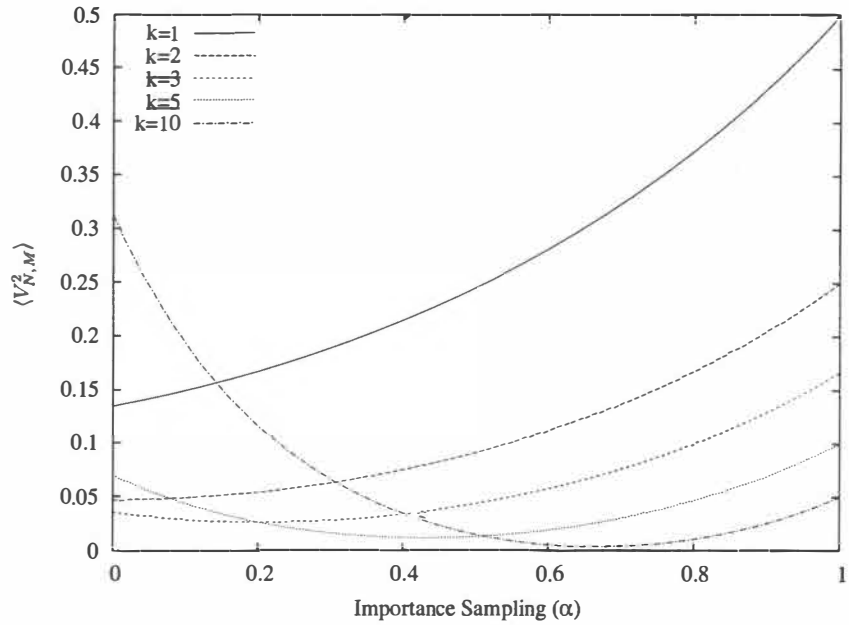


Figure 4.18: Variance of the $\mathbf{x}^2$ operator vs importance sampling (α) for various dimensions ($k = 1, 2, 3, 5, 10$).

4.7 Variances of the $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ operator

Results for the operator $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ are similar to those of the $\mathbf{x}^2$ operator. Plotting ln $\langle V^2_{N,M} \rangle$ vs dimension (figure 4.19), we see that with increasing dimension, the variance again decreases to a minimum before increasing. For large k, the variance grows exponentially with dimension. The minimum variance for this operator occurs at larger values of dimension than for the $\mathbf{x}^2$ operator. Overall, the statistical variance of $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ is slightly larger than that of the $\mathbf{x}^2$ operator in all dimensions. As we turn on importance sampling, the variance is visibly suppressed with weaker dependence on dimension. Not surprisingly, values of dimension k and importance sampling α that minimize the statistical variance of the $\mathbf{x}^2$ operator do not minimize the variance of the $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ operator.

In figure 4.20, we have plotted the expected variance vs importance sampling for several different dimensionalities. Even with full importance sampling, there is a finite variance associated with this operator, just as for the $\mathbf{x}^2$ operator. A trend toward worse statistics with better importance sampling is quite pronounced at low dimensionalities. This same behavior was noted earlier for the $\mathbf{x}^2$ operator and seemed to improve in higher dimensions. Yet, even in ten dimensions, the minimum in the variance for this operator occurs near $\alpha = 0.4$—nowhere near perfect importance sampling. This poses some serious difficulties to the computation of observables in QMC calculations, since statistical errors for arbitrary observables may be much larger than previously suspected.

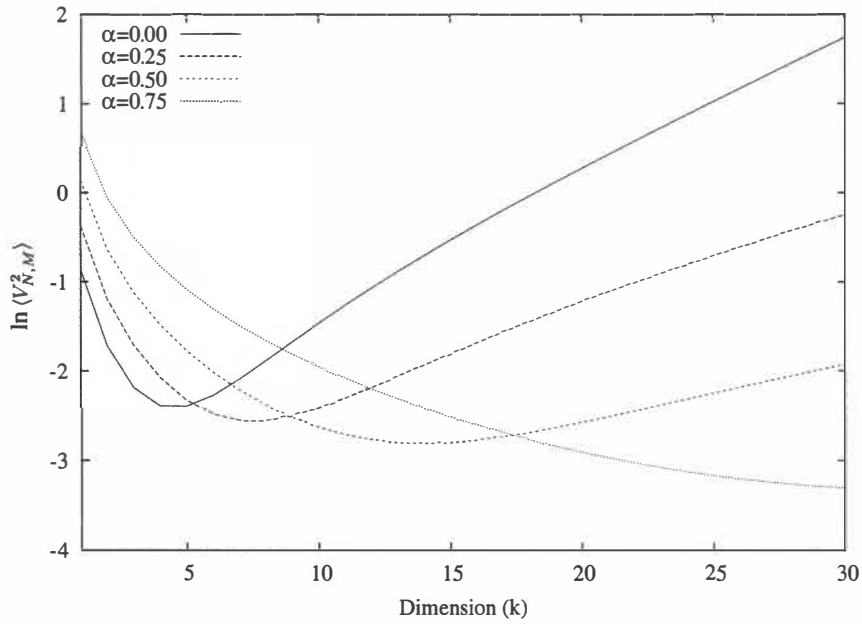


Figure 4.19: Natural logarithm of the variance of the $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ operator vs dimension (k) for various levels of importance sampling ($\alpha = 0.0, 0.25, 0.50, 0.75$).

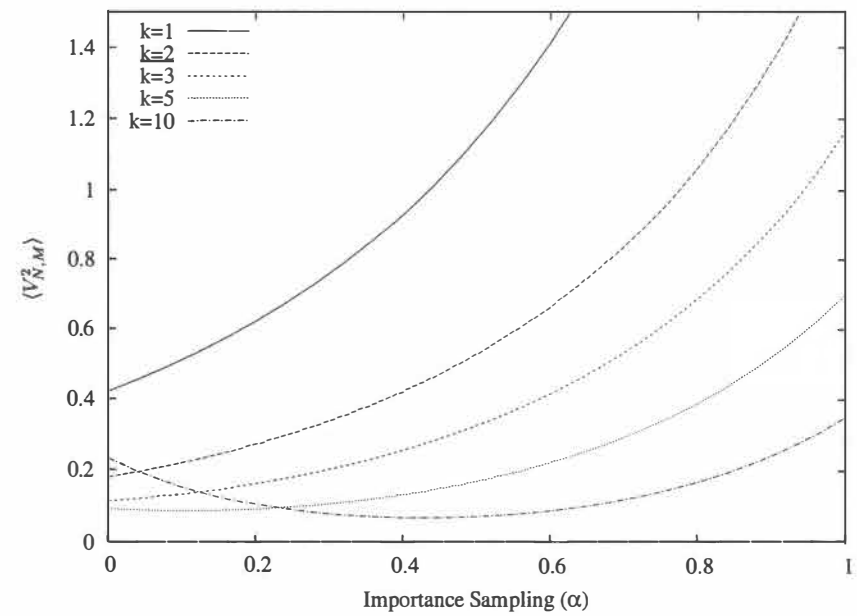


Figure 4.20: Variance of the $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ operator vs importance sampling (α) for various dimensions ($k = 1, 2, 3, 5, 10$).

4.8 Variances of the local energy operator

Of more practical interest is the local energy operator, $(\psi_T{}^{-1}\hat{\mathcal{H}}\psi_T)$. Since this operator is the primary source for quantum Monte Carlo estimates of the energy, it is important to understand its statistical error and how that error scales with dimension and importance sampling. For the harmonic oscillator trial wavefunction we have chosen (equation 4.3), the local energy operator is explicitly given as

$$\hat{E}_L = \frac{1}{2}\alpha - \frac{1}{2}\alpha^2\mathbf{x}^2 + \frac{1}{2}\mathbf{x}^2 \tag{4.7}$$

which contains a term proportional to the coefficient operator $(\hat{1})$ and other terms proportional to the $\mathbf{x}^2$ operator. Thus, the local energy operator is a linear combination of both the coefficient and $\mathbf{x}^2$ operators and should exhibit properties of both. This can be seen in figure 4.21. The variance curves have minima like those of the $\mathbf{x}^2$ operator. However, the location of the minima do not appear to change with dimension and are always located near $k = 3$. At large k, we see the same exponential behavior as for the coefficient and $\mathbf{x}^2$ operators.

As we increase importance sampling the variance decreases in all dimensions to zero (figure 4.22). Note that in equation 4.7 as $\alpha \to 1$, the local energy becomes constant and the dependence on $\mathbf{x}^2$ vanishes. In this limit we also expect the variance to behave like the variance of the coefficient operator and decrease to zero for all dimensions. When the exact ground state wavefunction is chosen as the trial wavefunction, there will be no statistical variation in the local

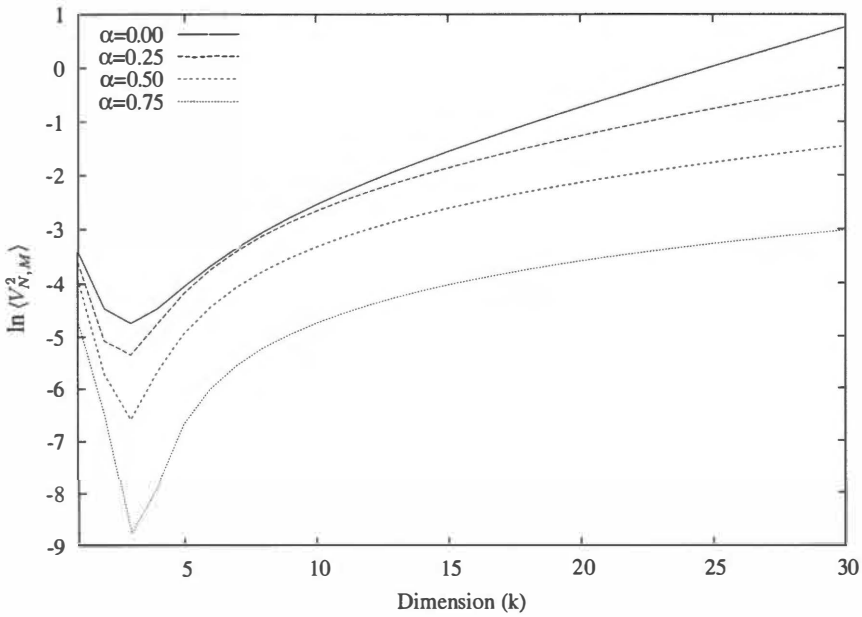


Figure 4.21: Natural logarithm of the variance of the local energy operator $(\psi_T{}^{-1}\hat{\mathcal{H}}\psi_T)$ vs dimension (k) for various levels of importance sampling $(\alpha = 0.0, 0.25, 0.50, 0.75)$.

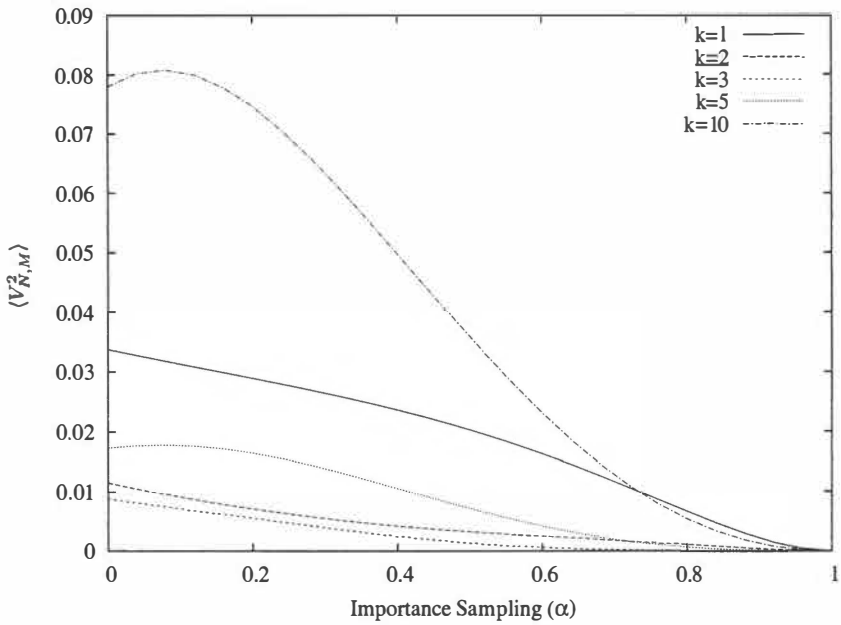


Figure 4.22: Variance of the local energy operator (${\psi_T}^{-1}\hat{\mathcal{H}}\psi_T$) vs importance sampling (α) for various dimensions ($k = 1, 2, 3, 5, 10$).

energy. This is a direct consequence of the trial wavefunction being an eigenfunction of $\hat{\mathcal{H}}$ when $\alpha = 1$. However, unlike the coefficient operator, the variance of $\hat{E}_L$ does not decrease monotonically to zero (as in the case $k = 10$). This is a result of the $\mathbf{x}^2$ character of the $\hat{E}_L$ operator. Unlike the coefficient operator, constant dimension curves appear to cross for the $\hat{E}_L$ operator. Additionally, statistical variances near zero are only encountered in the neighborhood of $\alpha = 1$. The slopes of the variance curves in this region tell us how sensitive the variance of the local energy is to the quality of the trial wavefunction.

4.9 Variance of the $(\hat{E}_L - \langle E_L \rangle)^2$ operator

Lastly, we consider the local energy variance operator, $(\hat{E}_L - \langle E_L \rangle)^2$, which is important in the variational optimization of trial wavefunctions. For a perfect trial wavefunction ($\alpha = 1$), it is known from variational theory that this operator has a zero expectation value. From figure 4.23 we see that the expected value of this operator does indeed go to zero for all states (n) as $\alpha \to 1$. For various levels of importance sampling, we note that the variance of this operator also goes through a minimum before growing exponentially (figure 4.24). Like the coefficient variance operator, this variance minimum is dependent on dimension, shifting toward higher dimension and broadening with increasing α. When we look at the variance as a function of the importance sampling parameter α (figure 4.25), it is evident that the variance of the operator $(\hat{E}_L - \langle E_L \rangle)^2$ also goes to zero in a

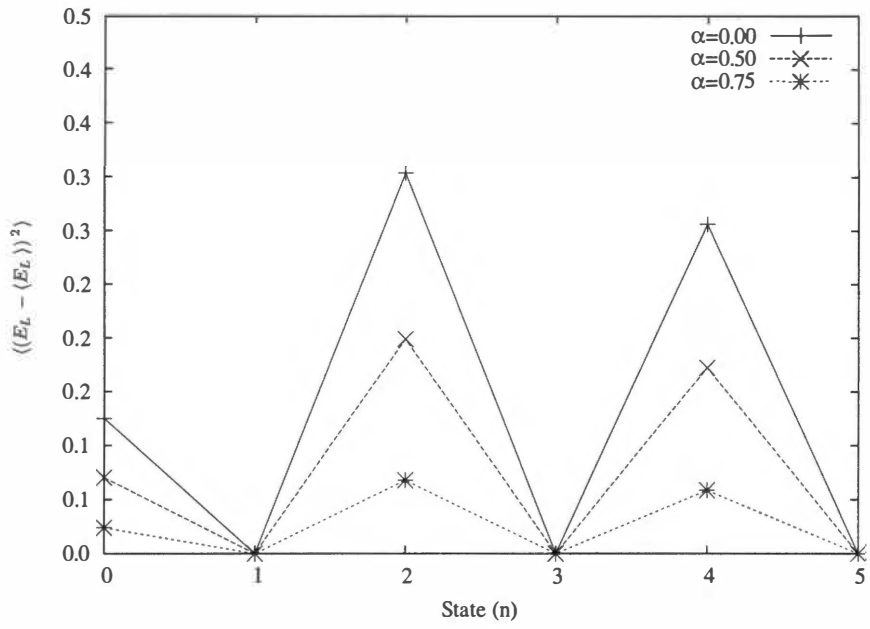


Figure 4.23: Expected value of the $(\hat{E}_L - \langle E_L \rangle)^2$ operator vs state (n) for various levels of importance sampling $(\alpha = 0.0, 0.50, 0.75)$.

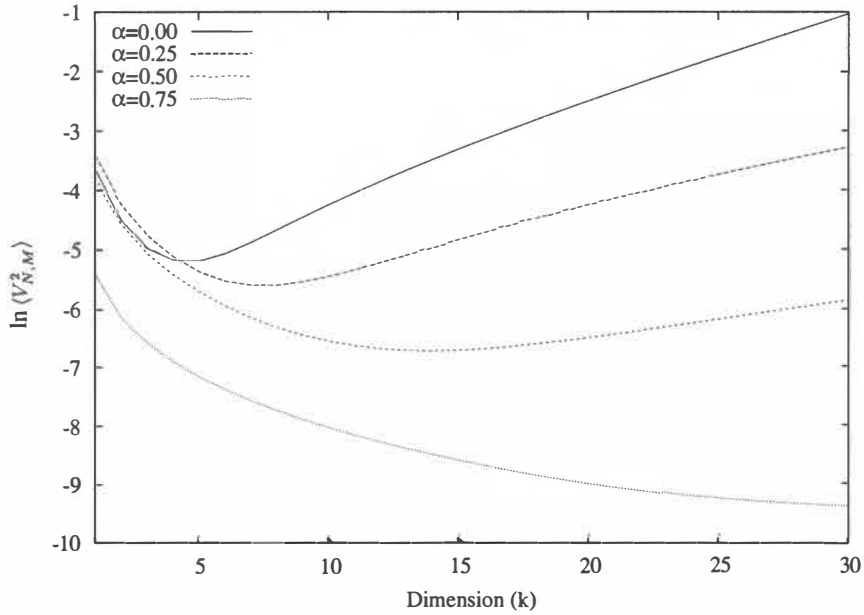


Figure 4.24: Natural logarithm of the variance of the $(\hat{E}_L - \langle E_L \rangle)^2$ operator vs dimension (k) for various levels of importance sampling $(\alpha = 0.0, 0.25, 0.50, 0.75)$.

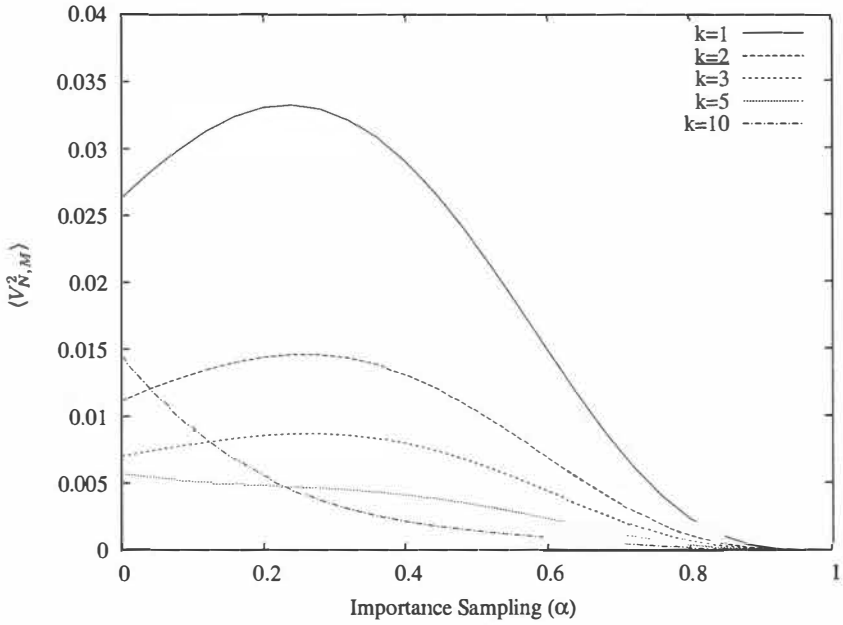


Figure 4.25: Variance of the $(\hat{E}_L - \langle E_L \rangle)^2$ operator vs importance sampling (α) for various dimensions ($k = 1, 2, 3, 5, 10$).

manner similar to the variance of the local energy operator.

4.10 Summary of results

In this chapter, we have considered the properties of the coefficient, $\mathbf{x}^2$, and $\hat{E}_L$ operators with respect to the harmonic oscillator. The coefficient operator obeys a zero-variance principle that causes the statistical variance to tend toward zero as $\alpha \to 1$. The $\mathbf{x}^2$ operator does not obey such a principle and has a finite statistical variance even when $\alpha = 1$. The local energy operator for the harmonic oscillator can be expressed as a linear combination of the coefficient operator and the $\mathbf{x}^2$ operator. As $\alpha \to 1$, the $\mathbf{x}^2$ character of the local energy operator disappears, forcing $\hat{E}_L$ to be proportional to the coefficient operator—a known zero-variance operator. Thus, the local energy also exhibits a zero-variance property in the limit of exact importance sampling.

We have also considered the properties of two quantum mechanical variance operators, $(\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2$ and $(\hat{E}_L - \langle E_L \rangle)^2$. The raw expectation values of these operators correspond to quantum mechanical observables. For the one dimensional harmonic oscillator ground state, $\langle \mathbf{x}^2 \rangle = 1/2$ and the variance, $\langle (\mathbf{x}^2 - \langle \mathbf{x}^2 \rangle)^2 \rangle$, equals

$$\left\langle \left(\langle \mathbf{x}^2 - \langle \mathbf{x}^2 \rangle \right)^2 \right\rangle = \langle \mathbf{x}^4 - \langle \mathbf{x}^2 \rangle \rangle = \frac{3}{4} - \left(\frac{1}{2} \right)^2 = \frac{1}{2}. \tag{4.8}$$

This result has no dependence on the sampling population (N) and is related in no way to statistical sampling. This expectation value is a property of the system. In contrast, the statistical variance from a quantum Monte Carlo simulation decreases with increasing population and cannot be taken as an estimate of $\langle(\mathbf{x}^2 - \langle\mathbf{x}^2\rangle)^2\rangle$. The same is true for the statistical variance of the local energy and the $(\hat{E}_L - \langle E_L\rangle)^2$ operator. Only in the case of unit population size ($N = 1$) and perfect importance sampling ($\alpha = 1$) are the two quantities equal. This means that both the observable variance and the corresponding statistical variance will follow similar zero-variance principles. Operators that possesses a zero-variance property have zero statistical variance at $\alpha = 1$, independent of the population size N. For operators that do not obey a zero-variance principle, it appears that an *effective* zero-variance principle may exist in high dimensional space, as described in chapter 3. For the operators $\mathbf{x}^2$ and $(\mathbf{x}^2 - \langle\mathbf{x}^2\rangle)^2$ examined here, importance sampling in the neighborhood of $\alpha = 1$ appears to minimize their variances in high dimensional systems.

As described above, the operators $\mathbf{x}^2$ and $\langle(\mathbf{x}^2 - \langle\mathbf{x}^2\rangle)^2\rangle$ do not possess a zero variance property. Additionally, values of α that minimize the variance for one operator do not minimize the variance of the other. This is an indication that a single importance sampling function is inadequate for controlling the variance of all observables in a QMC calculation. Separate or auxiliary importance sampling functions tuned for specific operators must be introduced to simultaneously minimize the variance of multiple observables. Such auxiliary functions must be optimized or parameterized to minimize the variance of each observable. This is the approach of Assaraf and Caffarel.[53,63,64] It is conceivable that a hybrid scheme employing both auxiliary importance sampling and forward walking could lead to the best general approach for reducing the statistical variance of arbitrary observables.

Chapter 5

Wavefunctions and Optimization

5.1 Introduction

Throughout the theoretical characterization of quantum Monte Carlo methods in the preceding chapters, we have largely ignored details of the trial wavefunction, Ψ_T. In this chapter, we consider the selection, optimization and implementation of trial wavefunctions suitable for ground-state calculations of helium and doped helium clusters. We begin with a survey of commonly used functional forms for cluster trial wavefunctions and proceed to develop an *ansatz* suitable for the study of doped helum clusters. Next we describe robust variance and energy minimization techniques for the optimization of trial wavefunctions. We conclude by reporting optimized parameters and VMC energies for energy-optimized ground state wavefunctions in He_n, Ca-He_n and Mg-He_n clusters for $n \leq 10$ and $n = 20$.

5.2 Ansatz for helium clusters

For van der Waals clusters of atoms it is common practice to employ a pair-product description in which the total wavefunction is a product of functions ϕ_{ij}, that depend only on a single internuclear separation r_{ij},

$$\Psi_T = \prod_{i<j} \phi(r_{ij}). \tag{5.1}$$

In helium clusters, this pair-product description provides a reasonably accurate and compact description of the total wavefunction, recovering over 90% of the total energy for small clusters. For higher accuracy calculations, the addition of three-body terms can lower the variational energy to within 2% of the total energy.[65] For many applications this improved accuracy is unnecessary since the DMC algorithm samples the exact ground state regardless of the quality of the trial wavefunction. One caveat is that properties computed using mixed or second-order estimators will still weakly depend on the quality of the trial wavefunction. Of course, the better the trial wavefunction, the lower the statistical error in estimates of the energy. Our goal of obtaining accurate estimates of the energy

and other properties of doped helium clusters drives our search for quality wavefunctions. For large clusters, outside the regime studied here, the inclusion of three-body terms is very beneficial and can substantially improve the quality of the trial wavefunction, albeit at additional computational cost. For the system and cluster sizes in which we are interested, well optimized pair-product forms are expected to be sufficiently accurate for DMC investigations.

Commonly used pair functions typically incorporate the Jastrow term[66]

$$\exp\left(\frac{p_5}{r_{ij}^v}\right), \tag{5.2}$$

to account for short-ranged repulsive interactions. The value $v = 5$ is chosen to satisfy cusp conditions for the Lennard-Jones potential, but v can be treated as a free parameter for potentials that differ in functional form.[65] The asymptotic pair wavefunction has the form

$$\exp\left(a_0 \ln(r_{ij}) + a_1 r_{ij}^b\right) \tag{5.3}$$

where b is near unity. Assembling these pieces into a more complete *ansatz*, including additional powers of r_{ij}^{-1}, we have

$$\phi_{ij} = \exp\left(a_0 \ln(r_{ij}) + a_1 r_{ij}^b + \sum_{v=0}^{5} a_v r_{ij}^{-v}\right) \tag{5.4}$$

which is the trial wavefunction employed by Whaley and collaborators[2,24,67] in calculations on helium, hydrogen, mixed hydrogen-helium clusters, respectively. The simplified *ansatz*

$$\phi_{ij} = \exp\left(\, p_0 \ln(r_{ij}) + p_1 r_{ij} + p_2 r_{ij}^{-2} + p_5 r_{ij}^{-5} \,\right), \tag{5.5}$$

retaining just the terms in the sum proportional to r_{ij}^{-5} and r_{ij}^{-2}, has also been successfully employed[21,27,68] in work with helium and noble gas clusters and by Mella *et al.*[16] in work with magnesium-doped helium clusters. However, as noted by Barnett and Whaley,[24] the r_{ij}^{-5} dependence in the repulsive region arising from the Lennard-Jones potential is not necessarily optimal for other types of potentials, such as the HFD-B form.[69] They instead employed the *ansatz*

$$\phi_{ij} = \exp\left(b_0 \ln(r_{ij}) + b_1 r_{ij} + e^{-\alpha r_{ij} - \beta r_{ij}^2} \sum_{k=0}^{5} t_k r_{ij}^k\right), \tag{5.6}$$

which incorporates a polynomial-exponential product to describe the short-range behavior. For potentials of the HFD-B form, this *ansatz* appears to yield a superior description of the wavefunction at short distances relative to the Jastrow form. A combination of the Jastrow form with a two parameter exponential term has been employed by Chiesa[68] and Casalegno,[70] also with good results. In this work we will employ the *ansatz* of Barnett and Whaley truncated to fewer terms as

$$\phi_{ij} = \exp\left(a_0 \ln(r_{ij}) + a_1 r_{ij}^b + \left(t_0 + t_1 r_{ij} + t_2 r_{ij}^2\right) e^{-\alpha r_{ij} - \beta r_{ij}^2}\right) \tag{5.7}$$

for both helium-helium and helium-dopant interactions. We find that optimized trial wavefunctions of this form yield up to 90% of the exact energy for all the pure and doped helium clusters we consider. For He_2, our optimized wavefunctions recover over 99% of the binding energy for all three He-He potentials employed in this work.

5.3 Trial wavefunction optimization

Before a trial wavefunction can be employed in a production-level calculation, it must first be *optimized* by generating an appropriate parameter set. A set of parameters $\{c_i\}$ (or equivalently a vector $\mathbf{c}$ in parameter space) is chosen such that the wavefunction minimizes the local energy or statistical variance in the local energy. In the limit of an exact trial wavefunction, both quantities will be minimized with the same parameter set. However, since all practical trial wavefunctions are only approximations to the exact form, energy and variance minimization procedures will generally yield different optimized parameters. Ideally, with a well chosen functional form, these differences will be small.

It has been suggested[71–74] that energy minimization may result in a slightly "better" wavefunction than variance minimization since it emphasizes the global accuracy of the wavefunction. Were it not for problems of numerical instability, energy minimization in VMC would be our first choice for trial wavefunction optimization. Unlike the variance of the energy which has a lower bound of zero and is always positive, the energy itself is not guaranteed to be bounded with respect to variations in the parameters when using a finite sample.[35,75] Consequently, attempting to minimize the energy of a fixed ensemble can result in unphysical values for parameters. This is not to say that energy minimization in VMC is impossible. With a reasonable set of starting parameters and a sampling population of generous size, reliable energy minimization is possible. However, for initial parameter sets corresponding to trial functions of unknown quality, variance minimization appears to be a more robust choice. Therefore, in an attempt to produce the best possible optimized wavefunctions, we will employ both energy and variance minimization techniques. Reliable variance minimization will be used to facilitate initial convergence to a good parameter set. Energy minimization will develop a well optimized wavefunction starting from the variance-optimized parameters. In this way, we obtain good energy-optimized wavefunctions while avoiding common numerical instabilities.

The full optimization of large parameter sets, particularly those containing non-linear parameters, can be difficult using VMC methods. Statistical uncertainties not present in standard deterministic minimization, often frustrate convergence and result in "false" minima.[71] In an attempt to overcome the problem of statistical noise, various minimization schemes have been developed, with the most successful method being *fixed ensemble minimization.* As the name suggests, deterministic minimization is performed with respect to a fixed or static sampling population. Since the sampling points do not change during the course of the minimization, convergence is usually not problematic. However, from ensemble to ensemble, the locations of the minima do differ, resulting in a range of parameter values. This is a consequence of finite sampling. Typically, the ensembles are

iterated a few times to obtain a reasonable parameter set. Realistically, a large number of ensembles would yield a statistical distribution of values for each parameter, from which the mean value and uncertainty could be estimated. Yet, this does not appear to be standard practice. In our work, we will use fixed parameter minimization and estimate parameter values from the minimization of many statistically independent ensembles.

Correlated sampling is essential to any VMC minimization. Since we are looking for changes in the local energy with respect to variations in the parameters, it is necessary to compute expectation values of the local energy

$$\left\langle \frac{\hat{\mathcal{H}}\Psi(\mathbf{x},\mathbf{c})}{\Psi(\mathbf{x},\mathbf{c})} \right\rangle = \left\langle \Psi(\mathbf{x},\mathbf{c}) \left| \frac{\hat{\mathcal{H}}\Psi(\mathbf{x},\mathbf{c})}{\Psi(\mathbf{x},\mathbf{c})} \right| \Psi(\mathbf{x},\mathbf{c}) \right\rangle \Big/ \left\langle \Psi(\mathbf{x},\mathbf{c}) \,\middle|\, \Psi(\mathbf{x},\mathbf{c}) \right\rangle \tag{5.8}$$

over the density $|\Psi(\mathbf{x},\mathbf{c})|^2$ for a variety of reparameterized wavefunctions. Unfortunately, in a standard VMC calculation, we have only constructed a Markov chain for the original wavefunction, which yields the density $|\Psi(\mathbf{x},\mathbf{c}_0)|^2$. Correlated sampling introduces a weight proportional to the ratio $|\Psi(\mathbf{x},\mathbf{c})/\Psi(\mathbf{x},\mathbf{c}_0)|^2$ so that

$$\langle E_L(\mathbf{x},\mathbf{c})\rangle = \left\langle \Psi(\mathbf{x},\mathbf{c}_0) \left| \frac{\Psi^2(\mathbf{x},\mathbf{c})}{\Psi^2(\mathbf{x},\mathbf{c}_0)} \frac{\hat{\mathcal{H}}\Psi(\mathbf{x},\mathbf{c})}{\Psi(\mathbf{x},\mathbf{c})} \right| \Psi(\mathbf{x},\mathbf{c}_0) \right\rangle \Big/ \left\langle \Psi(\mathbf{x},\mathbf{c}_0) \left| \frac{\Psi^2(\mathbf{x},\mathbf{c})}{\Psi^2(\mathbf{x},\mathbf{c}_0)} \right| \Psi(\mathbf{x},\mathbf{c}_0) \right\rangle \tag{5.9}$$

and we can now compute the above expectation value using readily available information. Of course, both wavefunctions must be evaluated at the same points, which requires that we compute both quantities from the same random walk.

5.3.1 Variance minimization

In this work we employ the variance minimization scheme of Umrigar,[76] which has become a relatively standard approach to the optimization of trial functions incorporating large parameter sets or a significant number of non-linear parameters. In variance minimization, we seek to optimize a set of parameters with respect to fluctuations in the local energy, $\Psi_T^{-1}\hat{\mathcal{H}}\Psi_T$. Instead of the actual variance of this operator (which would take an entire simulation to compute accurately), we employ the second moment of the local energy about a fixed reference value, E_R, according to

$$\tilde{\sigma}^2(E_L) = \left(\frac{\hat{\mathcal{H}}\Psi_T(\mathbf{x},\mathbf{c})}{\Psi_T(\mathbf{x},\mathbf{c})} - E_R \right)^2 . \tag{5.10}$$

Generally, E_R is chosen to approximate the true ground state energy and can be estimated from a DMC run using a crudely optimized wavefunction. As the wavefunction is improved, more reliable of estimates of E_R may be employed.

To perform the minimization, we select an independent set of configurations from $\Psi_T(\mathbf{x},\mathbf{c}_0)$ that we take to be fixed during the minimization procedure. Within this fixed ensemble, we then minimize the second moment of the local energy with respect to all parameters in the parameter set.

Correlated sampling, as described above, is used to accurately compute differences in the variance between parameterized versions of $\Psi_T(\mathbf{x}, \mathbf{c})$. At this point, one would typically update the initial parameter set $\mathbf{c}$ and repeat the optimization procedure. However, there are several problems with this approach. If the ensemble size is too small, the "optimized" parameter set may not be a good estimate of the best possible parameter set. This is because it is relatively easy for unphysical sets of parameters to make the variance $\tilde{\sigma}^2(E_L)$ zero or nearly zero in small sampling populations. Moreover, if there are multiple minima, it is also possible that the optimization procedure will falsely converge to a local minimum instead of the global minimum. Such convergence would depend largely on the initial values of the parameters.

To overcome these issues, we use a modified approach. Instead of updating the current parameter set, we draw another (independent) ensemble from $\Psi_T(\mathbf{x}, \mathbf{c}_0)$ and repeat the minimization procedure to find a new estimate of the optimized parameter set. This is necessary because a single fixed ensemble by itself is not representative of the entire set of such ensembles. After drawing many such ensembles, we will obtain a statistical distribution of optimized parameter values. If all optimized parameter sets correspond to the same minimum, we would simply average over all parameters to determine the "best" set.

More generally, the resulting parameter sets may correspond to several different minima. In this case, averaging over all optimal parameter sets would provide a poor estimate of the "best" parameter set. In such a case, we must seek to differentiate parameter sets belonging to different minima, extract only the parameter sets that correspond to the lowest minimum, and average over the reduced pool of parameter estimates. The selection of minima can be performed visually by looking at cross-sectional scatter plots for each ordered pair of parameters, (c_j, c_k). Parameters corresponding to the same minimum will be clustered together. Figures 5.1 and 5.2 are representative scatter plots for the optimization of He_2. As parameters are refined from iteration to iteration, we see both a shift in parameter space and a reduction in the scatter. The data cluster with the lowest average variance is taken to correspond to the global minimum. Parameters corresponding to other minimia, if present, are then culled from the data set. This provides a fairly robust prescription for zeroing in on the correct minimum.

This procedure is also useful for determining the topology of parameter space. Tight clustering of parameters indicates a strong dependence of the quality of the wavefunction on the parameter set. Likewise, diffuse clustering can indicate a weak dependence. Often, parameters will exhibit a certain degree of dependence on other parameters. This is indicated by highly elliptical clustering and can be profitable in identifying superfluous or redundant parameters in the *ansatz*.

Once we are assured that all remaining parameter estimates correspond to the minimum of interest, we may average over the parameters to determine the "best" set. We then update the initial wavefunction using these new parameters and repeat the process. Generally, we adjust the convergence thresholds in the optimization procedure to yield more accurate parameter estimates for the second and third iterations. This procedure results in a good variance-optimized variational wavefunction that will be our starting point for the energy minimization procedure.

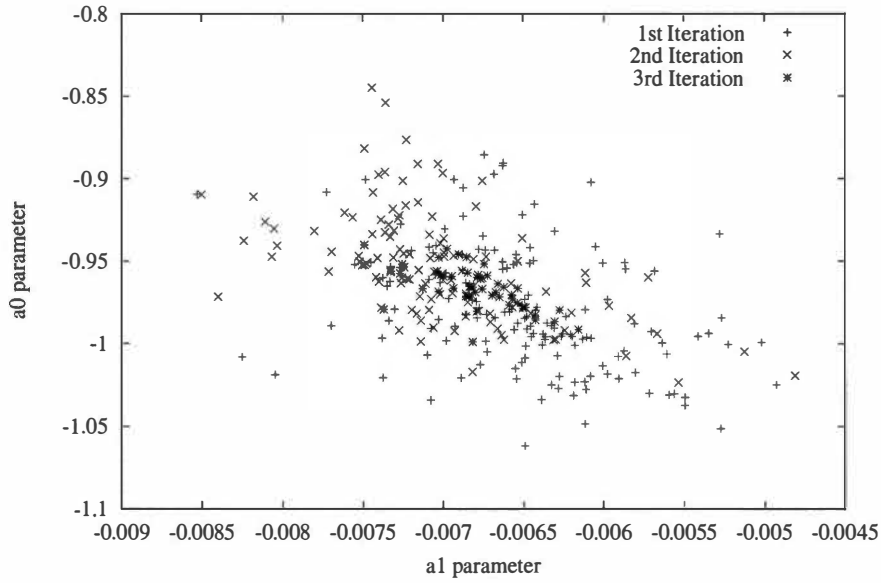


Figure 5.1: Cross-sectional scatter plot for the a_1 and a_0 parameters of the He_2 trial wavefunction using the HFD-B(87) potential.

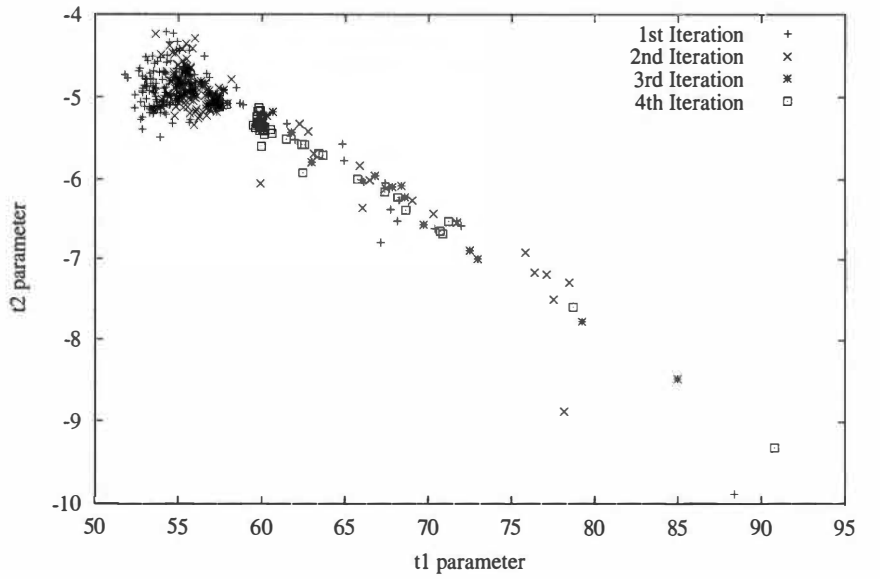


Figure 5.2: Cross-sectional scatter plot for the t_1 and t_2 parameters of the He_2 trial wavefunction using the HFD-B(87) potential.

5.3.2 Energy minimization

While variance minimization techniques have become a rather universal approach to the optimization of trial wavefunctions using VMC, energy minimization techniques have been largely ignored. This is largely because variance minimization exhibits better numerical stability than energy minimization and because the variance has a known lower bound of zero. The difficulty in minimizing the energy stems from the fact that the energy lacks boundedness with respect to variations in the parameters.[35,75] Even in a finite sample, we are always guaranteed to obtain a positive variance. This is not the case with the energy. As a result, the optimization procedure can "walk off" into regions of parameter space where the energy is increasingly negative. Recently, Liu, Zhang, and Rappe[73] have employed analytical first and second energy derivatives in an effort to produce a stable and reliable energy minimization scheme. A practical alternative is to start the energy optimization process near the desired minimum and to use a large sampling population. Making use of a variance optimized initial wavefunction, the first criteria is easy to meet. The second criteria is required to practically limit the "unboundedness" of the energy. Since for small sampling populations the local energy distribution is skewed (usually toward lower energies), it is important to limit excursions into these tails. For a large enough sampling population, the local energy distribution will become a Gaussian. In this limit, excursions beyond three standard deviations are relatively rare occurrences. Together, these two criteria vastly improve the robustness of energy minimization within the context of VMC.

The general scheme for energy minimization largely parallels that of our variance minimization procedure. We select fixed ensembles from an appropriate density and optimize wavefunction parameters using correlated sampling to minimize the local energy. The starting parameters are taken to be the optimal set produced by the variance minimization procedure described above. Unwanted excursions are very unlikely but still possible when starting with a good set of parameters. To prevent these, we employ a maximum number of iterations or steps within the optimization algorithm instead of enforcing a convergence cutoff. Generally the maximum number of steps is large enough to allow the optimization to converge normally but small enough to stop an unbounded walk. As before, large sampling populations are employed and many different fixed ensembles are minimized to generate a collection of optimized parameter vectors. Unbounded excursions will show up as outliers on the 2D scatter plots and can be effectively culled from the data set before averages are taken.

In some cases, it may be advantageous to limit the optimization step size and reduce the maximum number of allowable steps. The smaller initial step size encourages the optimization procedure to fall toward the nearest minimum. The use of fewer allowable steps causes the optimization to move toward the minimum, but not necessarily reach it. Both strategies are important because they help to limit the expenditure of computational effort on the optimization of unphysical parameter sets.

5.3.3 Justification for modified optimization prodedures

The theoretical justification for our modified procedures arises out of the work of Bressanini and coworkers.[71] They observe for various test cases that the local energy distribution is strikingly non-Gaussian. Viewed as a fitting procedure based on a Gaussian model, the optimization process will then unfairly weight the tails of this distribution. To deal with this problem we must either choose another maximum likelihood model or or work with quantities that do obey Gaussian statistics.

Consider the local energy distribution defined as

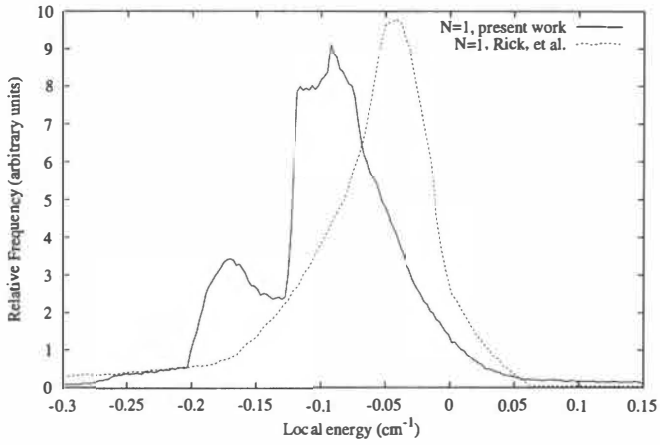
$$\mathcal{E}(E) = \int \Psi_T(\mathbf{x})\delta(E - E_L(\mathbf{x}))\Psi_T(\mathbf{x})d\mathbf{x}. \tag{5.11}$$

For a specific value of E, the delta function picks out all configurations yielding that energy. In the limit $\Psi_T \rightarrow \Psi_0$, all configurations will generate the same energy and the local energy distribution will be a delta function located at E_0. Since each energy is estimated using only one configuration, this is equivalent to sampling with the population size $N = 1$. When ensemble sizes greater than one are employed, it is important to consider the distribution

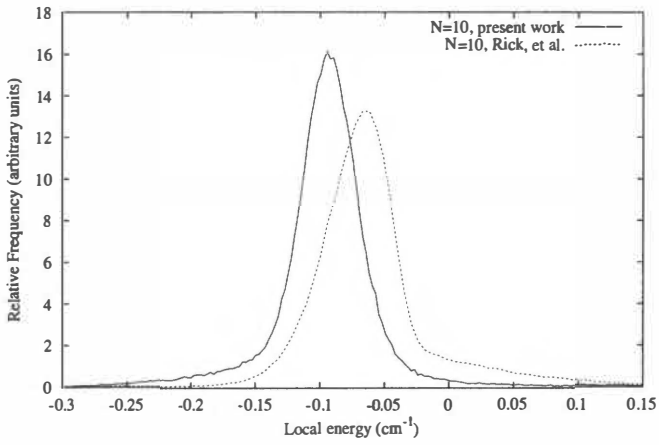
$$\bar{\mathcal{E}}_N(E) = \int \Psi_T(\mathbf{x})\delta(E - \bar{E}_{L,N}(\mathbf{x}))\Psi_T(\mathbf{x})d^N\mathbf{x} \tag{5.12}$$

where $\bar{E}_L(\mathbf{x})$ is the average local energy over N configurations and the integral is now taken over N dimensions. As we increase the population size, the central limit theorem guarantees that the local energy distribution will evolve into a Gaussian distribution. Applied to the energy optimization process, we find that the use of a large sampling population, though expensive, vastly improves the robustness of the optimization without the necessity of having to invoke other maximum likelihood models.

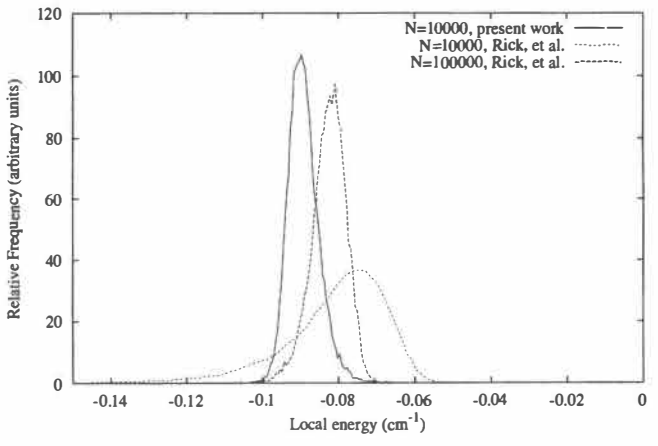
Several researchers have remarked on the skewed shape of the local energy distribution[27] for He_3 and how it affects the optimization process.[71] However, we find that the degree of skewness present in distributions of the local energy is largely dependent on the optimized form of the trial wavefunction. It is possible that the existence of such asymmetrical skewness directly correlates with a systematic deficiency in the functional form of the trial wavefunction. In any case, any non-Gaussian local energy distribution can be made Gaussian by using a large enough sampling population. Thus, significant deviations from a Gaussian local energy distribution directly indicate that the size of the sampling population is inadequate. This is most clearly demonstrated in the population dependence of the local energy distributions for He_3 illustrated in figure 5.3. As we increase the population size, the local energy distributions narrow up and become more Gaussian-like. At larger population sizes (panel (c)), the quality of the present trial function becomes readily apparent, yielding a significantly lower variational energy than the commonly employed[16,27,71] trial function of Rick *et al.*[21] We also note that the population size necessary to obtain a comparable statistical error for the present trial wavefunction is an order of magnitude smaller than required



(a)



(b)



(c)

Figure 5.3: Local energy distributions for He_3 comparing the present optimized trial function with that of Rick *et al.* for populations sizes of (a) $N = 1$, (b) $N = 10$, and (c) $N = 10000$.

for the trial wavefunction of Rick *et al.* This will directly benefit the energy optimization process since we can achieve Gaussian statistics with smaller sampling populations and thereby improve the numerical stability of the algorithm. This also underscores the importance of choosing a trial functions with sufficient functional flexibility. Trial wavefunction with insufficient flexibility lead to greater statistical error and demand larger population sizes to maintain a robust optimization procedure.

5.3.4 Optimization algorithms

There are a variety of optimization algorithms that may be employed in the fixed ensemble minimization procedure. Typically, conjugate gradient[77] and line minimization[77] algorithms are employed since they provide the most general and efficient approaches to multi-dimensional minimization. For energy minimization, more sophisticated methods incorporating analytical derivatives are available. However, these methods may not always be the most advantageous choices for energy minimization since the use of gradients can exacerbate the aforementioned numerical instabilities. Instead, we choose to use the more crude, but highly robust simplex method.[77] For the numerical optimization of the relatively small eight parameter wavefunctions used here, the added reliability far outweighs the inefficiency of the method.

5.4 Interatomic potentials

5.4.1 Neglect of many-body effects

In the study of atomic clusters, it is common practice to break down the many-body potential into the summation

$$V(\mathbf{x}) = \sum_{i<j} V_{i,j} + \sum_{i<j<k} V_{i,j,k} + \ldots, \tag{5.13}$$

where the $V_{i,j}$ and $V_{i,j,k}$ are two and three body interaction terms, respectively. In many cases, the two body terms are sufficient to capture the essential physics. In the case of helium, which has a small polarizability, the contribution of three body terms is frequently neglected. Specifically for magnesium doped helium clusters, Mella[16] has verified the assumption of pairwise additivity, finding three-body effects to only be important at very close range and at small He-Mg-He bond angles. Thus, we will completely ignore three-body effects in the present work and focus solely on the two-body interaction potentials.

5.4.2 Helium-helium potentials

A variety of interatomic helium pair potentials are available in the literature for the accurate modeling of helium-helium interactions. In this work, we will primarily make use of the recent SAPT2-retarded(97) potential of Janzen and Aziz.[78] For comparison with previous quantum Monte Carlo studies of helium clusters, we will also employ the HFD-B(87) potential of Aziz, McCourt

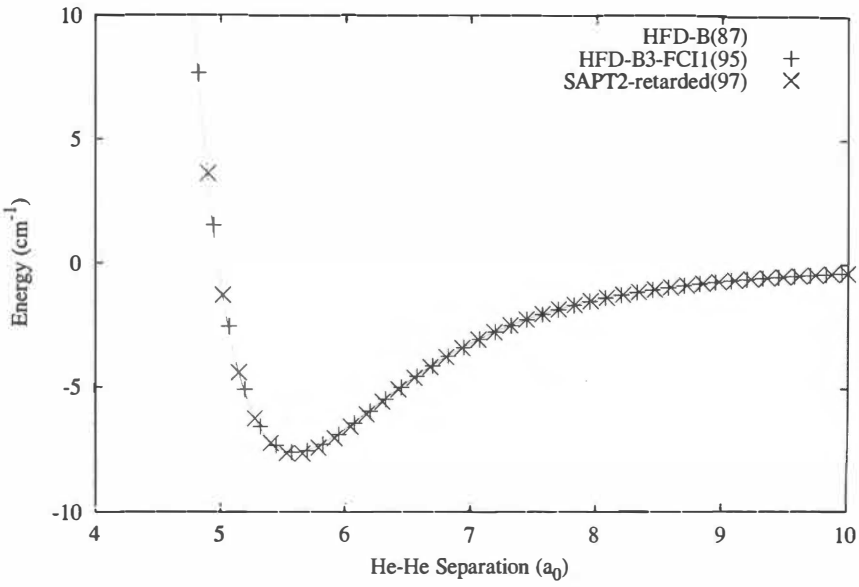


Figure 5.4: He-He potentials employed in this work.

and Wong.[69] Additionally, we consider the recent HFD-B3-FCI1(95)[79] form which demonstrates weaker attractions than the other two potentials that we have selected. Though not employed here, other notable helium potentials include the Tang-Toennies-Yiu (TTY) potential,[80] the FCI/CBS potential of van Mourik and Dunning,[81] and the exact quantum Monte Carlo results of Anderson.[82] The potential curves for the HFD-B(87), HFD-B3-FCI1(95), and SAPT2-retarded(97) potentials are plotted in figure 5.4. Since the differences between these potentials are difficult to see, we have also plotted the differences, relative to the HFD-B(87) potential in figure 5.5.

HFD-B(87) potential

The HFD-B(87) potential, though somewhat dated, is surprisingly accurate and has been frequently employed in the investigation of helium systems. It is based on the Hartree-Fock plus damped dispersion model with the repulsive wall being fit to the Green's function Monte Carlo results of Ceperley and Partridge.[83] The HFD-B(87) potential has a well depth of 7.609 cm^{-1} and supports a single bound state with energy 1.1546x10^{-3} cm^{-1}. Functionally, the reduced form of the HFD-B(87) potential is

$$V_{HFD-B(87)} = \epsilon \left(A \exp(-\alpha x + \beta x^2) - F(x) \sum_{j=0}^{2} \frac{c_{2j+6}}{x^{2j+6}} \right) \tag{5.14}$$

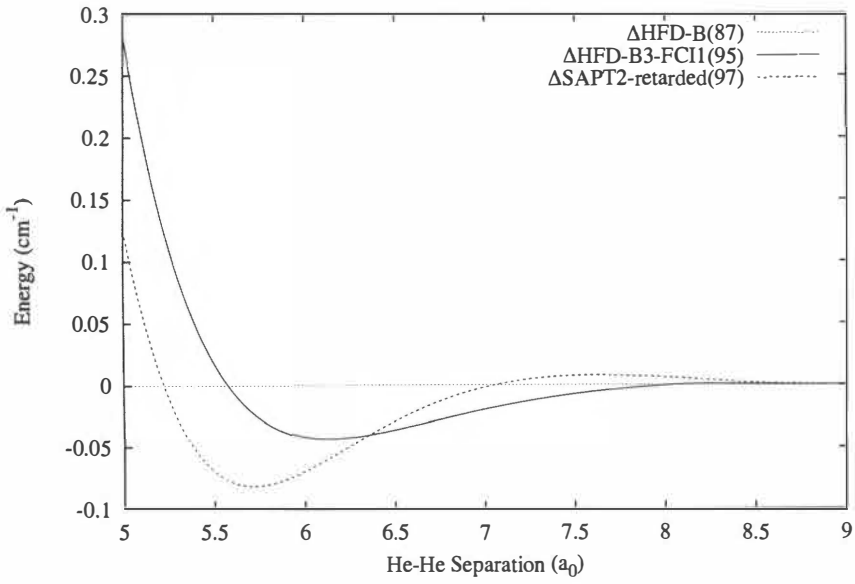


Figure 5.5: Differences in He-He potentials relative to the HFD-B(87) potential.

where

$$F(x) = \begin{cases} \exp\left[-\left(D/x - 1\right)^2\right] & (x < D) \\ 1 & (x \geq D) \end{cases} \tag{5.15}$$

and $x = r/r_m$. Numerical values for the parameters can be found in table 5.1. We have chosen to use the HFD-B(87) potential in our work primarily as a benchmark for comparison with other studies but also as a standard for judging the effect of small differences in the helium-helium potential.

HFD-B3-FCI1(95) potential

The more recent HFD-B3-FCI1(95) potential of Aziz, Janzen, and Moldover is a reparameterization (table 5.2) of the original HFD-B(87) form, incorporating the more recent full configuration interaction (FCI) results of van Mourik and van Lenthe.[84] The HFD-B3-FCI1(95) potential, like the HFD-B(87) version, incorporates the Green's function Monte Carlo results of Ceperley and Partridge[83] at short range. It has a well depth of 7.615 cm^{-1} (slightly deeper than the HFD-B(87) version) and supports a single bound state with binding energy 1.0875x10^{-3} cm^{-1}, slightly weaker bound than that of the HFD-B(87) potential. This weaker binding, similar to that of the TTY potential (9.1467e-4 cm^{-1}), leads to noticeable energetic differences in larger clusters.

Table 5.1: Parameters for the HFD-B(87) potential.

Parameter	Value
A	1.8443101e5
α	10.43329537
c_6	1.36745214
c_8	0.42123807
c_{10}	0.17473318
β	-2.27965105
D	1.4826
ϵ	7.609 cm^{-1}
r_m	5.599 a_0

Table 5.2: Parameters for the HFD-B3-FCI1(95) potential.

Parameter	Value
A	1.86924404e5
α	10.5717543
c_6	1.35186623
c_8	0.41495143
c_{10}	0.17151143
β	-2.07758779
D	1.438
ϵ	7.615 cm^{-1}
r_m	5.609 a_0

SAPT2-retarded(97) potential

The SAPT2 helium-helium potential proposed by Janzen and Aziz[78] is the most refined of the three helium-helium potentials employed in this study. The potential is fit to a modified Tang-Toennies form[85,85] and incorporates the infinite order symmetry adapted perturbation theory calculations of Korona *et al.*[86] at intermediate range, the Green's function Monte Carlo results of Ceperley and Partridge[83] at close range, and the dispersion coefficients of Bishop and Pipin.[87] Relativistic retardation effects are applied to the dipole-dipole dispersion terms over the range of 5.7 to 10^5 bohr. The retarded SAPT2 potential has a well depth of 7.688 cm^{-1}, appreciably deeper than either of the HFD-B potentials. The binding energy of 1.2476x10^{-3} cm^{-1} is also greater than that of the HFD-B potentials. Consequently, we expect the SAPT2 potential to yield noticeably lower energies for the larger helium clusters in this study. The functional form of the SAPT2 potential is given as

$$\begin{aligned} V_{SAPT2} &= A\, e^{-\alpha r+\beta r^2} - \left[1 - \left(\sum_{k=0}^{6}(\delta r)^k/k!\right) e^{-\delta r}\right] C_6 f(r)/r^6 \\ &\quad - \sum_{n=4}^{8}\left[1 - \left(\sum_{k=0}^{2n}(\delta r)^k/k!\right) e^{-\delta r}\right] C_{2n}/r^{2n} \end{aligned} \tag{5.16}$$

where $f(r)$ is a position dependent retardation coefficient described by polynomials in r (see Janzen and Aziz[78]). The parameters for the SAPT2-retarded potential can be found in table 5.3.

Table 5.3: Parameters for the SAPT2-retarded(97) potential.

Parameter	Value
A	6.56912828 E_h
α	1.88648251 $bohr^{-1}$
β	-6.20013490x10^{-2} $bohr^{-2}$
δ	1.94861295 $bohr^{-1}$
C_6	1.4609778 a.u.
C_8	14.117855 a.u.
C_{10}	183.69125 a.u.
C_{12}	3.265x10^3 a.u.
C_{14}	7.644x10^4 a.u.
C_{16}	2.275x10^6 a.u.
ϵ	7.68789 cm^{-1}
r_m	5.60234 bohr

5.4.3 Metal-helium potentials

Magnesium-helium potential

To describe the interactions between magnesium dopants and helium atoms, we use the pair potential of Hinde[14] which incorporates a full treatment of coupled-cluster triple excitations (CCSDT) and core-valence correlation corrections. Large, nearly saturated sets of bond functions are used to recover a large percentage of the correlation energy. Consequently, we expect this potential to perform slightly better than the CCSD(T) potential of Partridge, Stallcop, and Levin[15] which only incorporates perturbative triples in the coupled-cluster expansion. To implement the potential in our quantum Monte Carlo investigation, we analytically extend the repulsive region of the curve using a simple exponential fit. At long range, we employ an averaging procedure to smoothly transition from the *ab initio* data to a pure 6-8-10 dispersion function. Dispersion coefficients computed from oscillator strength sum rules (C_6 = 21.45 a.u., C_8 = 0.8864e3 a.u., C_{10} = 3.936e4 a.u.) are taken from the work of Mitroy and Bromley[88] and are in good agreement with those of Standard and Certain.[89] A cubic spline is used to interpolate between computed *ab initio* points.

The resulting potential yields a well depth of 5.00 cm^{-1}, an r_e value of 5.10 Å, and a ground state energy of -0.8884 cm^{-1} as determined by a direct numerical solution of the one dimensional Schrödinger equation. For comparison, the CCSD(T) potential of Partridge *et al.* gives a well depth of 4.76 cm^{-1}, an r_e value of 5.09 Å, and a ground state energy of -0.791175 cm^{-1} when extended with the C_6 and C_8 coefficients of Mitroy and Bromley.[88] Lovallo and Klobukowski[90] have also developed a CCSD(T) magnesium-helium potential using a model core potential (MCP) approach. They report similar parameters (ϵ = 4.69 cm^{-1}, r_e = 5.11 Å) and an energy of -0.7686 cm^{-1}, close to the value of Partridge *et al.* Funk *et al.*[91] have also computed an MP4 surface (ϵ = 4.54 cm^{-1}, r_e = 5.16 Å, E_0 = 0.734 cm^{-1}) which is significantly shallower than the recent MP4 calculations of Mella[16] (ϵ = 5.70 cm^{-1}, r_e = 5.03 Å). For all QMC and bound state results, we use the nuclear mass of ^{24}Mg, 23.9784589 amu, taken from NIST data. The CCSDT(Hinde) Mg-He potential is plotted in figure 5.6 along with the potentials of Partridge *et al.* and Funk *et al.* for comparison.

Calcium-helium potential

We have chosen to model the calcium-helium interactions also with the *ab initio* potential of Hinde[14] (figure 5.7). Like the magnesium-helium potential described above, the calcium-helium results were derived using an identical CCSDT approach employing CCSD(T) core-valence correlation corrections and a saturated set of bond functions. The potential parameters (ϵ = 3.43 cm^{-1}, r_e = 5.9) are in good agreement with those of Partridge *et al.*[15] (ϵ = 3.31 cm^{-1}, r_e = 5.95) but differ slightly from those of Lovallo and Klobukowski[90] (6.04, 3.14 cm^{-1}). As before, we extend the *ab initio* data using an exponential fit to the repulsive region and a dispersion fit at long distances to the coefficients of Mitroy and Bromley[88], (C_6 = 36.59 a.u., C_8 = 2.138e3 a.u., C_{10} = 13.19e4 a.u.). Here, both the C_6 and C_8 coefficients lie outside the range established by Standard and Certain.[89] Numerical integration with Hinde's potential yields a ground state of -0.570239 cm^{-1}, slightly lower than Lovallo and Klobukowski's value of -0.4793 cm^{-1} which employs the dispersion

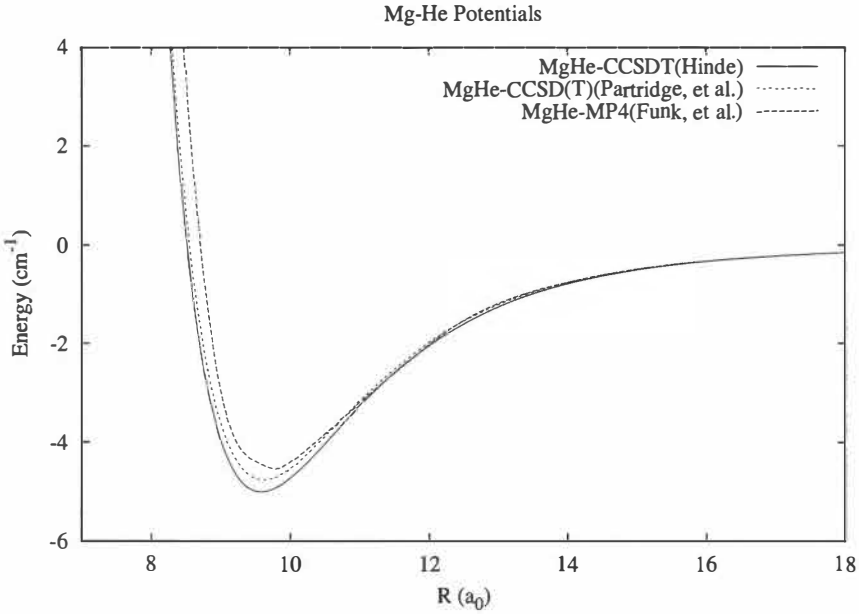


Figure 5.6: Mg-He CCSDT(Hinde) potential.

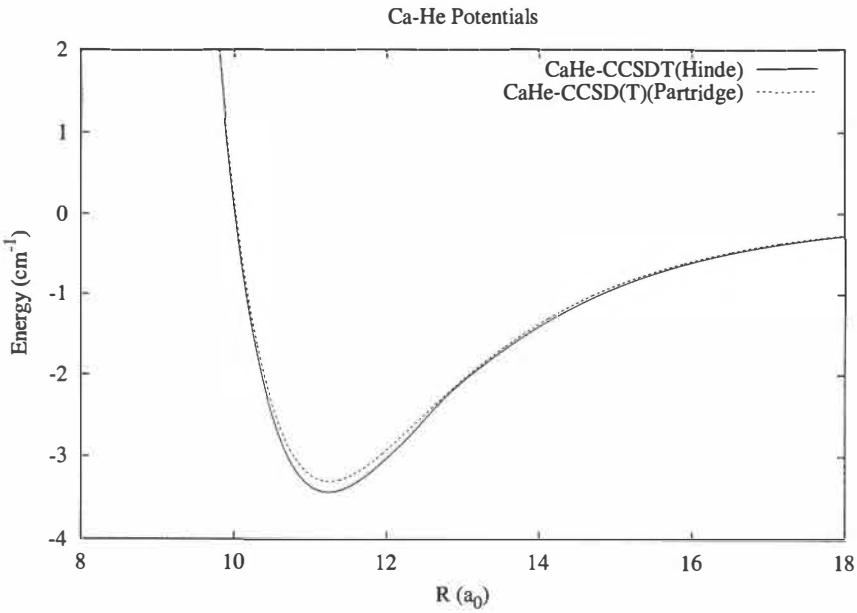


Figure 5.7: Ca-He CCSDT(Hinde) potential.

coefficients of Standard and Certain. The bound state energy of -0.52891 cm^{-1} obtained from the potential of Partridge *et al.* using Mitroy and Bromley's C_6 and C_8 coefficients, agrees even better with the present estimate. In all calculations, we take the nuclear mass of ^{40}Ca to be 39.9516196 amu. Figure 5.7 shows a plot of the Ca-He CCSDT(Hinde) potential as well at that of Partridge *et al.*

5.5 Optimized cluster wavefunctions

5.5.1 Pure helium clusters

Wavefunctions for pure helium clusters were optimized using the *ansatz*

$$\phi_{ij} = \exp\left(a_0 \ln(r_{ij}) + a_1 r_{ij}^b + \left(t_0 + t_1 r_{ij} + t_2 r_{ij}^2\right) e^{-\alpha r_{ij} - \beta r_{ij}^2}\right). \tag{5.17}$$

We employ a "bootstrap" approach beginning with a fit to the numerical solution of the 1D Schrödinger equation. The parameters are then optimized within VMC, first with a variance minimization procedure, followed by energy minimization. After convergence, the fully optimized trial wavefunction is then used as the starting point for optimizing the next cluster having an additional helium atom. This process is repeated to generate optimized trial functions for helium clusters of size $n = 2-10$ and $n = 20$. We generate three such sets of optimized wavefunctions, one set for each of the three helium potentials that we have selected. The optimized parameters for the HFD-B(87) potential can be found in table B.1, the parameters for the HFD-B3-FCI1(95) potential in table B.2 and the parameters for the SAPT2-retarded(97) potential in table B.3. Here, and throughout this work, we have chosen to report our statistical errors at 95% confidence (denoted by brackets[ ]) and defined as

$$result \pm 1.960 \frac{s}{\sqrt{M}} \tag{5.18}$$

where s is the sample standard deviation and M is the number of blocks. Note that this differs from the common literature practice of reporting the standard error ($s/\sqrt{M}$) (denoted by parenthesis ()) or standard deviation s of a result.

The optimized wavefunctions are of excellent quality. In table 5.4, we compare the the variational total energies to corresponding DMC calculations (chaper 6) for the present *ansatz* and for the calculations of Lewerenz.[27] For all cluster sizes, we find that the present *ansatz* recovers at least 2% more of the total cluster energy than does the Rick *et al.*[21] trial wavefunction employed by Lewerenz. For He_3, the results are more dramatic, with the present trial wavefunction recovering nearly 8% more of the total energy. We also see a smooth trend toward poorer wavefunctions as we move to larger cluster sizes. Since our trial function includes no explicit flexibility for describing additional solvation structure beyond nearest neighbor interactions, it is no surprise that it performs gradually worse with increasing cluster size. However, for the relatively compact 8 parameter trial function employed here, the results are quite good and surprisingly improved over trial forms frequently employed in the literature.

Table 5.4: Quality of present helium wavefunctions judged by the VMC recovery of the total DMC energy (table 6.1) using the HFD-B(87) potential.

	Lewerenz[27]		Present	
n	VMC	% of DMC	VMC	% of DMC
2	—	—	-0.001146[1]	99.28
3	-0.0802(16)	88.13	-0.08941[2]	96.83
4	-0.3706(28)	92.37	-0.38969[3]	96.50
5	-0.871(1)	93.86	-0.8843[2]	95.42
6	-1.533(1)	93.08	-1.5643[3]	95.01
7	-2.357(2)	93.05	-2.3947[4]	94.44
8	-3.291(3)	92.24	-3.3527[5]	93.83
9	-4.331(3)	91.55	-4.4299[7]	93.44
10	-5.502(5)	91.47	-5.6040[8]	92.92
20	—	—	-21.032[2]	89.71

5.5.2 Magnesium doped helium clusters

Magnesium-doped helium cluster wavefunctions were energy optimized in a manner similar to the pure helium cluster wavefunctions. For the doped clusters, we use the CCSDT potential of Hinde for the Mg-He interactions and the SAPT2-retarded(97) potential for the He-He interactions. We do not expect the choice of He-He potential to significantly affect energetic or structural properties since differences in the CCSDT(Hinde) and CCSD(T)(Partridge *et al.*) Mg-He potentials are much larger and lead to noticeably distinct energies even in the ground state of the Mg-He dimer. Another difference in the case of the optimization of magnesium doped helium clusters is that we choose to optimize both the Mg-He wavefunction and the He-He wavefunction *simultaneously*. Since the presence of an impurity alters the compactness of the He moiety we find it useful to take this into account during our optimization process.

The optimized parameters for the Mg-He and He-He wavefunctions are listed in table B.4 and table B.5, respectively. Note that there is a clear difference between the He-He wavefunctions for pure and doped clusters, much larger than the differences between pure cluster wavefunctions optimized with the three different He-He potentials. This indicates the non-negligible influence of the magnesium dopant on the structure of the helium cluster.

Comparing the VMC energy to the corresponding DMC energy for each cluster size, (table 5.5) we find that the resulting wavefunctions are also of excellent quality. At all cluster sizes, we recover a larger percentage of the DMC energy in the doped clusters as compared to the pure helium clusters. For $MgHe_{20}$, we recover nearly 92% of the total energy, which attests to the high quality of the present trial wavefunction form and its applicability in describing cluster impurities.

Table 5.5: Quality of $MgHe_n$ wavefunctions judged by the VMC recovery of the total DMC energy (table 6.5) using the SAPT2-retarded(97) potential.

n	VMC	% of DMC
1	-0.888373[4]	~100.00
2	-1.91745[2]	99.24
3	-3.07862[5]	98.37
4	-4.35874[7]	97.55
5	-5.7421[1]	96.78
6	-7.2143[2]	96.07
7	-8.7647[3]	95.50
8	-10.3830[4]	95.00
9	-12.0618[5]	94.52
10	-13.7957[5]	94.13
20	-33.420[2]	91.69

5.5.3 Calcium doped helium clusters

Wavefunctions for the calcium doped clusters were energy optimized in exactly the same way as for the magnesium doped clusters. As before, both the He-He and He-dopant wavefunctions were optimized simultaneously to ensure the best possible parameter sets for the chosen functional forms. For the Ca-He interaction potential, we employ the CCSDT potential of Hinde extended with the dispersion coefficients of Mitroy and Bromley[88]. Again, the helium-helium interactions are described the SAPT2-retarded(97) potential of Janzen and Aziz.[78]

The optimized trial wavefunction coefficients are collected in tables B.6 and B.7. Like the optimized wavefunctions for magnesium doped clusters, we see considerable modifications to the He-He trial functions with the calcium dopant present. This again attests to the non-negligible influence of these alkaline earth metal dopants on the structure of the helium cluster, unlike for alkali metals.

Repeating our quality analysis of the optimized trial functions (table 5.6), we find that the optimized trial functions for calcium doped clusters are also of high quality, recovering over 90% of the DMC energy for all cluster sizes studied. The good variational energies for both pure and doped clusters further confirms the wide applicability of our chosen *ansatz*.

5.5.4 Summary

In this chapter, we have described a novel optimization scheme for the accurate determination of trial wavefunction parameters within the context of variational Monte Carlo. While computa-

Table 5.6: Quality of $CaHe_n$ wavefunctions judged by the VMC recovery of the total DMC energy (table 6.4) using the SAPT2-retarded(97) potential.

n	VMC	% of DMC
1	-0.570207[1]	~100.00
2	-1.26011[2]	99.12
3	-2.08185[4]	98.07
4	-3.04002[7]	96.95
5	-4.1273[3]	96.08
6	-5.3279[5]	95.35
7	-6.6302[7]	94.76
8	-8.023[1]	94.39
9	-9.492[1]	93.79
10	-11.034[2]	93.45
20	-29.012[4]	90.03

tionally intensive, the method appears robust with respect to the optimization of trial wavefunction in weakly bound cluster systems. Using this new optimization technique, we have obtained well optimized wavefunctions for pure and alkaline earth metal doped helium clusters. This is the first such trial wavefunction *ansatz* optimized specifically for the study of these special doped helium systems, particularly with regard to the simultaneous optimization of both dopant-helium and helium-helium trial functions. Consequently, we predict that these new optimization techniques and high quality wavefunctions will be extremely valuable for reducing the statistical error and bias in the present work and in future QMC investigations of doped helium systems.

Chapter 6

Solvation Features of Mg and Ca Doped Helium Clusters

6.1 Simulation details

In obtaining the results presented in this chapter, we have taken great care to assure the accuracy and reliability of the computed values. We use the specially optimized wavefunctions of the preceding chapter to greatly suppress the statistical errors and aid convergence in our DMC runs. In all cases, we use the largest computationally feasible equilibrium population sizes, nearly $5x10^5$ for the $MgHe_2$ and $CaHe_2$ clusters down to $1x10^5$ for the largest $MgHe_{20}$ and $CaHe_{20}$ clusters. Several values of the timestep were used down to 50 hartree^{-1} to ensure accurate extrapolation to $\Delta\tau = 0$. Simulation lengths anywhere from $2x10^4$ to $1x10^5$ steps were employed, with an effective block size of about 400 steps for $\tau = 50$. Because of the exceedingly large memory requirements associated with generating the distribution functions (particularly the two dimensional distributions), they were generated in a separate batch of runs employing less demanding simulation parameters. All simulations were performed using a custom "in house" suite of QMC codes developed and optimized specifically for the study of clusters. Computations were performed on a fifty-node Linux Beowulf cluster composed of 2.4GHz dual Xeon machines linked with a channel-bonded fast ethernet interconnect. The amount of memory per node ranged from one to three gigabytes. Typical single processor calculations requird approximately one week of cpu time.

6.2 Energetics

6.2.1 Total energies of pure He_n clusters

Before we examine the results for the doped clusters, we need to first consider the results for pure helium clusters, since these will be used to compute the binding energies of the magnesium and calcium dopants. We have computed the total DMC energies (table 6.1) as a function of increasing

Table 6.1: Total DMC energies (cm^{-1}) for pure He_n clusters for the HFD-B(87), HFD-B3-FCI1(95), and SAPT2-retarded(97) potentials. Brackets denote errors at 95% confidence.

n	HFD-B(87)	HFD-B3-FCI1(95)	SAPT2-retarded(97)
2	-0.001153[2]	-0.00107[3]	-0.00122[3]
3	-0.09234[4]	-0.09108[3]	-0.09379[3]
4	-0.4010[1]	-0.3975[1]	-0.4052[1]
5	-0.9267[2]	-0.9207[2]	-0.9347[2]
6	-1.6464[4]	-1.6370[3]	-1.6596[3]
7	-2.5358[8]	-2.5206[7]	-2.5556[7]
8	-3.573[1]	-3.5529[8]	-3.5976[9]
9	-4.741[2]	-4.716[2]	-4.775[2]
10	-6.031[2]	-6.004[2]	-6.068[2]
20	-23.443[7]	-23.317[7]	-23.574[8]

cluster size ($2 \leq n \leq 10, n = 20$) for the HFD-B(87), HFD-B3-FCI1(95), and SAPT2-retarded(97) potentials. The DMC energies for the helium dimers agree very well with results from direct numerical integration of the one dimensional Schrödinger equation for these three potentials which give $-0.00115459\ cm^{-1}$, $-0.00108752\ cm^{-1}$, and $-0.00124756\ cm^{-1}$, respectively. The results also agree very closely from potential to potential indicating that the energy is largely independent of the choice of He-He potential. Here, we have intentionally converged the DMC energy tightly enough to unambiguously determine the differences in cluster energies due to differences in the potentials. Depending on the potential, we see a 1% spread in the computed cluster energies. This relative uncertainty will also be present in the results for the doped clusters and should be taken into account when comparing with other results in the literature. Since the differences due to the choice of helium potential are slight, we choose only the newest SAPT2-retarded(97) potential to model helium-helium interactions in the doped systems.

The results for the HFD-B(87) potential are useful for comparison to other work and as a check on the validity of our coded algorithm. In table 6.2 we compare DMC energies computed in the present work to values obtained from the literature. We note that all reported values agree with those of Whaley[24,25] to within the stated error bars. In several cases, our values lie slightly outside the errorbars of Lewerenz[27] and lower in energy, particularly for He_3 and for the largest cluster sizes investigated. It is possible that the disagreement is a result of any number of subtle factors such as the choice of helium nuclear mass, which we take to be 4.00150609 amu. In light of the excellent DMC reproducibility of the numerically determined helium dimer energies and the close agreement with results of Whaley, we are convinced of the validity of our programmed DMC implementation.

Table 6.2: Comparison of total DMC energies (cm^{-1}) for pure He_n clusters computed using the HFD-B(87) potential. Brackets denote errors at 95% confidence and parentheses denote standard errors.

n	Lewerenz[27]	Whaley[24,25]	Present
3	-0.0910(5)	-0.0923(4)	-0.09234[4]
4	-0.4012(5)	-0.4017(8)	-0.4010[1]
5	-0.928(1)	-0.929(4)	-0.9267[2]
6	-1.647(1)	—	-1.6464[4]
7	-2.533(2)	-2.536(3)	-2.5358[8]
8	-3.568(2)	—	-3.573[1]
9	-4.731(4)	—	-4.741[2]
10	-6.015(5)	—	-6.031[2]
20	—	-23.46(3)	-23.443[7]

6.2.2 Evaporation energies of pure He_n clusters

From the DMC results for the total energies, we may compute the corresponding evaporation energies. The evaporation energy is defined as

$$E_{evap} = E(n-1) - E(n) \tag{6.1}$$

which is the energy required to remove a helium atom from a cluster containing n helium atoms. The computed evaporation energies for all the helium-helium potentials are listed in table 6.3 and are plotted in figure 6.1. The helium evaporation energies increase smoothly and monotonically with increasing cluster size showing no preferred stable or "magic" cluster sizes. This observed lack of atomistic structure is consistent with the unique fluid-like characteristics of these clusters. At $n = 20$ we are still far from the experimental bulk value of 4.95 cm^{-1}, indicating the likely persistence of finite size effects for a large range of cluster sizes. The smoothness of the curve is strong evidence that our results are well converged at the simulation parameters employed in the DMC calculations.

Again, we see a similar trend of the HFD-B3-FCI1(95) potential binding less strongly than the other two, giving consistently smaller values for the evaporation energies. Likewise, the SAPT2-retarded(97) potential appears to bind the strongest of the three potentials examined. However, the differences are still within 1% of one another.

6.2.3 Total energies of Ca- and Mg-doped He_n clusters

DMC energies for helium clusters doped with calcium and magnesium appear in tables 6.4 and 6.5, respectively. The results for calcium are the first reported calculations for this dopant. Con-

Table 6.3: DMC evaporation energies (cm^{-1}) for pure He_n clusters for the HFD-B(87), HFD-B3-FCI1(95), and SAPT2-retarded(97) potentials.

n	HFD-B(87)	HFD-B3-FCI1(95)	SAPT2-retarded(97)
3	0.0912[1]	0.0900[1]	0.0926[1]
4	0.3087[2]	0.3064[2]	0.3114[2]
5	0.5257[4]	0.5232[3]	0.5295[3]
6	0.7197[8]	0.7163[7]	0.7249[7]
7	0.889[1]	0.8836[8]	0.896[1]
8	1.037[2]	1.032[2]	1.042[2]
9	1.168[2]	1.163[2]	1.177[2]
10	1.290[7]	1.287[7]	1.293[8]

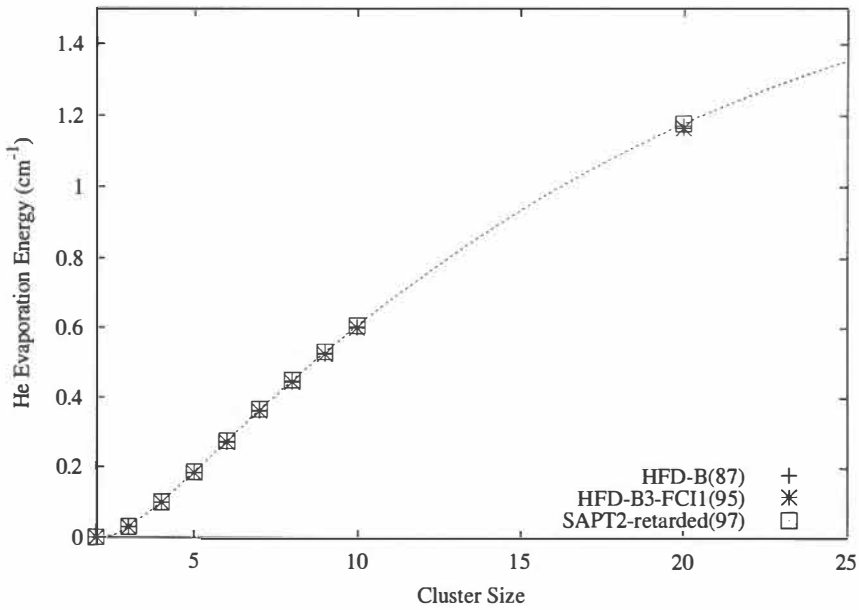


Figure 6.1: He_n evaporation energies for the HFD-B(87), HFD-B3-FCI1(95) and SAPT2-retarded(97) potentials as a function of cluster size. The curve is provided as a guide to the eye, but is physically motivated and representative of the expected trend.

Table 6.4: Total DMC energies (cm^{-1}) for doped $CaHe_n$ clusters computed using the SAPT2-retarded(97) and CCSDT(Hinde) potentials. Brackets denote errors at 95% confidence.

n	Present
2	-1.27129[7]
3	-2.1229[2]
4	-3.1357[4]
5	-4.2958[7]
6	-5.588[1]
7	-6.997[2]
8	-8.500[2]
9	-10.120[3]
10	-11.808[4]
20	-32.226[35]

Table 6.5: Total DMC energies (cm^{-1}) for doped $MgHe_n$ clusters computed using the SAPT2-retarded(97) and CCSDT(Hinde) potentials. Comparison is made to the results of Mella *et al.* computed with various potentials. Brackets denote errors at 95% confidence and parentheses denote standard errors.

n	Present CCSDT(Hinde)	MP4[16]	CCSD(T)[16]	CCSDT(Hinde)[16]
2	-1.9321[1]	-2.5366(6)		
3	-3.1298[3]			
4	-4.4682[5]	-5.660(2)	-4.127(4)	
5	-5.9333[8]			
6	-7.5092[13]	-9.2445(9)	-6.9881(9)	
7	-9.177[2]			
8	-10.929[3]	-13.172(5)	-10.244(5)	
9	-12.761[4]			
10	-14.656[4]			
20	-36.450[25]	-41.54(2)	-34.76(2)	-35.894(7)

sequently, there are no values in the literature available for comparison. The careful calibration and validation of our DMC algorithm for the case of pure helium clusters gives us confidence in the accuracy of the present values. Compared to the energies for pure helium, we find that the total energies are noticeably lower and decrease smoothly as a function of cluster size.

For magnesium, the total energies are slightly lower than those of calcium, indicating stronger binding of the magnesium dopant. Recent results of Mella *et al.*[16] are included in table 6.5 for comparison. Unfortunately, their computed values using the CCSDT(Hinde) potential overlap with the present work only for a cluster size of $n = 20$. Their result of -35.894(7) for $MgHe_{20}$ agrees quite well with our own estimate, -36.450[25]. The small disagreement is partly due to the differences in the analytical extension of the *ab initio* potential. Mella *et al.* employ only a C_6 extrapolation, while we use a more accurate C_6-C_8-C_{10} extrapolation with the recent dispersion coefficients of Mitroy and Bromley.[88] They also use the TTY potential for the helium pair interactions where presently, we have chosen to work with the SAPT2-retarded(97) potential. Since our work demonstrates a roughly 1% variation with the choice of helium-helium potential, it is not unreasonable to assume that this difference is wholly explainable by a combination of these two factors. Thus our fully independent calculations corroborate the accuracy of their work while at the same time providing additional proof that our implementation of the DMC algorithm is correct.

Though not directly comparable, the energies reported by Mella *et al.* for their CCSD(T) potential are slightly higher than for the CCSDT(Hinde) potential, but track ours reasonably well. Their MP4 DMC results come from a significantly deeper potential which we reproduce here to show the range of variation that follows from the use of different potential models.

6.2.4 Evaporation energies of Ca- and Mg-doped He_n clusters

As for the pure helium clusters, we can also compute the evaporation energies of helium in these doped clusters. Figure 6.2 compares the evaporation energy curve for pure helium clusters with that of calcium doped clusters. Like pure helium clusters, the trend is smooth and monotonically increasing with cluster size. No noticeable structure or preferred cluster sizes are observed. The actual evaporation energies are noted in table 6.6.

The helium evaporation energies for magnesium doped clusters are plotted in figure 6.3. The trend is again smooth, monotonically increasing and slightly higher in energy than the curve for calcium doped clusters. Also plotted are the results of Mella *et al.*[16] for their MP4 and CCSD(T) potentials. As expected, the MP4 potential binds more strongly and follows the trend of the present results. However, the reported results for the CCSD(T) potential are lower than the present results and appear to have a slightly different curvature. The evaporation energies may be found in table 6.7 and are compared with the MP4 and CCSD(T) DMC results of Mella *et al.*

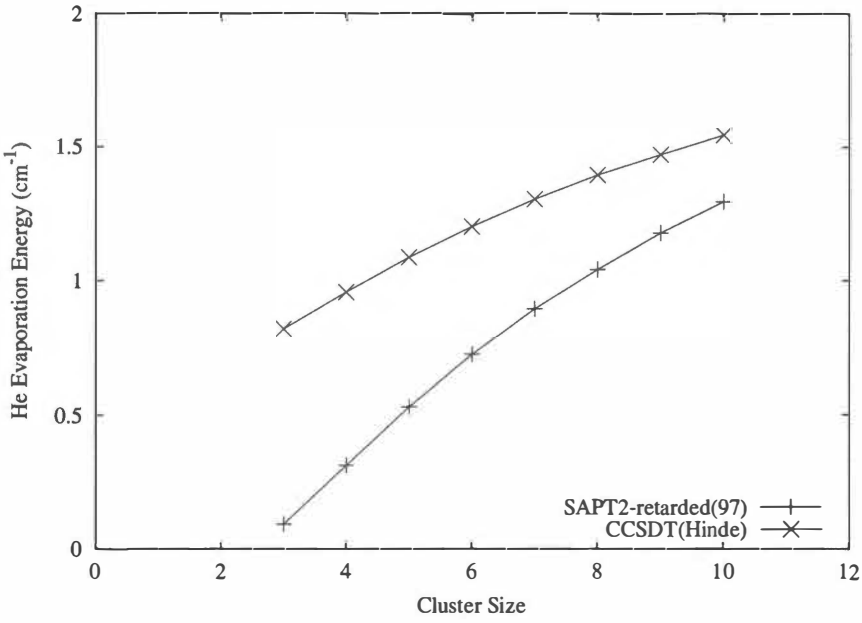


Figure 6.2: Helium evaporation energies in $CaHe_n$ as a function of cluster size n. Evaporation energies for pure helium clusters are provided for comparison.

Table 6.6: Helium evaporation energies (cm^{-1}) in doped $CaHe_n$ clusters as a function of cluster size n. Brackets denote errors at 95% confidence.

n	Present
3	0.82174[2]
4	0.9582[4]
5	1.0873[7]
6	1.201[1]
7	1.302[2]
8	1.393[2]
9	1.469[3]
10	1.543[4]

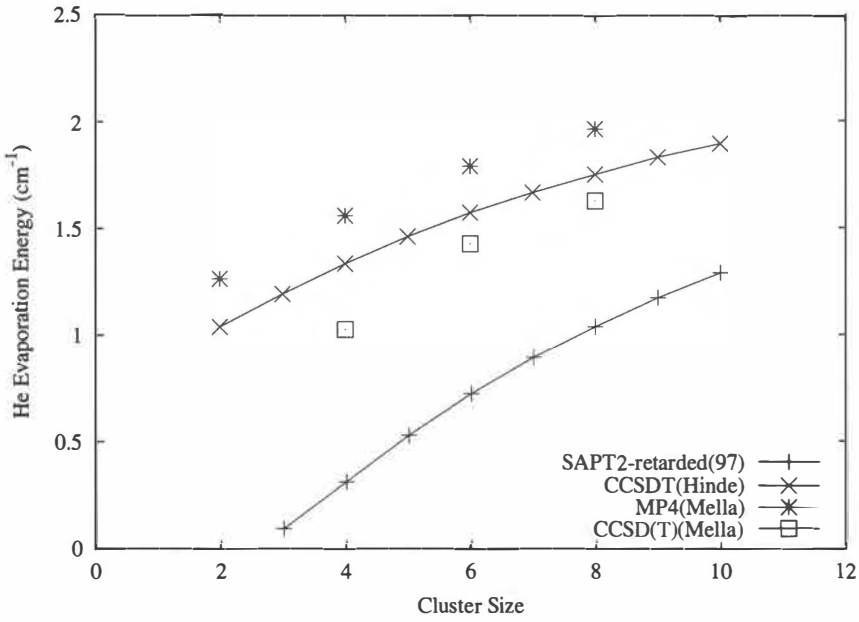


Figure 6.3: Helium evaporation energies (cm^{-1}) in $MgHe_n$ as a function of cluster size. Evaporation energies from the work of Mella *et al* and for pure helium clusters are provided for comparison.

Table 6.7: Helium evaporation energies (cm^{-1}) in doped $MgHe_n$ clusters as a function of cluster size. Brackets denote errors at 95% confidence and parentheses denote standard errors.

n	Present CCSDT(Hinde)	MP4[16]	CCSD(T)[16]
2	1.0437[1]	1.2683(3)	
3	1.1977[3]		
4	1.3384[5]	1.5618(8)	1.0317(4)
5	1.4651[8]		
6	1.576[1]	1.7921(9)	1.431(1)
7	1.668[2]		
8	1.752[3]	1.964(2)	1.628(3)
9	1.832[4]		
10	1.895[4]		

6.2.5 Dopant binding energies

Now that we have obtained total energies for both the pure and doped helium clusters, we can calculate the binding energy of dopants by subtracting the energy of the doped clusters from that of the pure clusters:

$$E_{bind} = E_{pure}(n) - E_{doped}(n). \tag{6.2}$$

Figure 6.4 plots the binding energy as function of cluster size for both calcium and magnesium dopants. Both curves are smooth and increasing with cluster size. At $n = 20$, the energy separation between the two dopants has grown to nearly 5 cm^{-1} and appears to still be increasing. The binding energies for calcium are tabulated in table 6.8 while those for magnesium appear in table 6.9 and are compared with the results of Mella *et al.*[16] In figure 6.5 we have again plotted the binding energy curve for magnesium for visual comparison with the curves of Mella *et al.* Compared with our results, we see characteristically similar behavior for their CCSD(T) potential, yielding slightly weaker binding energies than ours. Their corresponding MP4 curve however, demonstrates a remarkably different curvature, in agreement with the full interior solvation of magnesium that they observe uniquely for that potential.

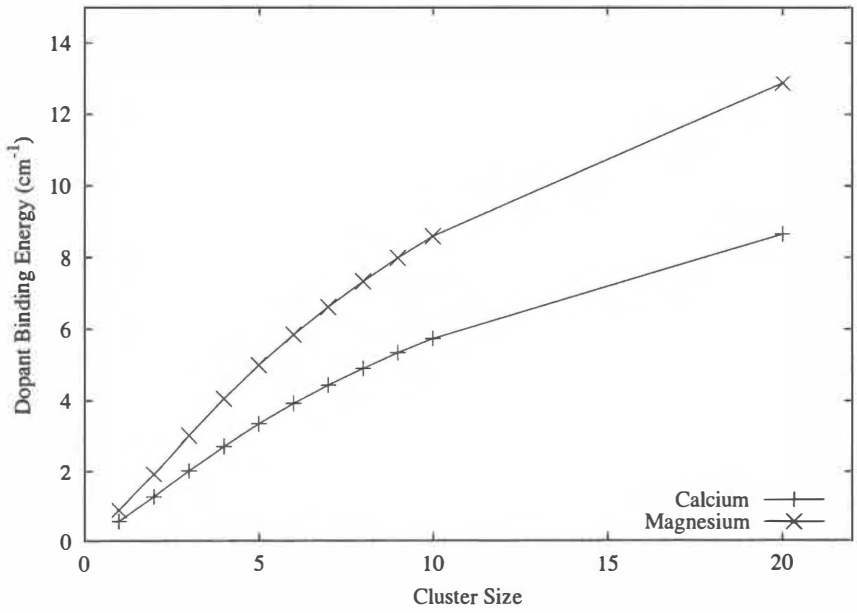


Figure 6.4: Comparison of calcium and magnesium binding energies (cm^{-1}) as a function of the cluster size (n).

Table 6.8: Calcium binding energies (cm^{-1}) as a function of cluster size. Brackets denote errors at 95% confidence.

n	Present
1	0.570239[a]
2	1.27004[7]
3	2.0291[2]
4	2.7305[4]
5	3.3611[7]
6	3.928[1]
7	4.441[2]
8	4.902[2]
9	5.345[3]
10	5.740[4]
20	8.652[35]

[a] Exact result

Table 6.9: Magnesium binding energies (cm^{-1}) as a function of cluster size. Brackets denote errors at 95% confidence and parentheses denote standard errors.

n	Present CCSDT(Hinde)	MP4[16]	CCSD(T)[16]	CCSD(T)(Hinde)[16]
1	0.888394[a]			
2	1.9308[1]	2.5357		
3	3.0360[3]			
4	4.0630[5]	5.271	3.738	
5	4.9986[8]			
6	5.850[1]	7.6368	5.3804	
7	6.622[2]			
8	7.332[3]	9.604	6.676	
9	7.986[4]			
10	8.588[4]			
20	12.877[25]	18.50	11.72	12.85

[a] Exact result

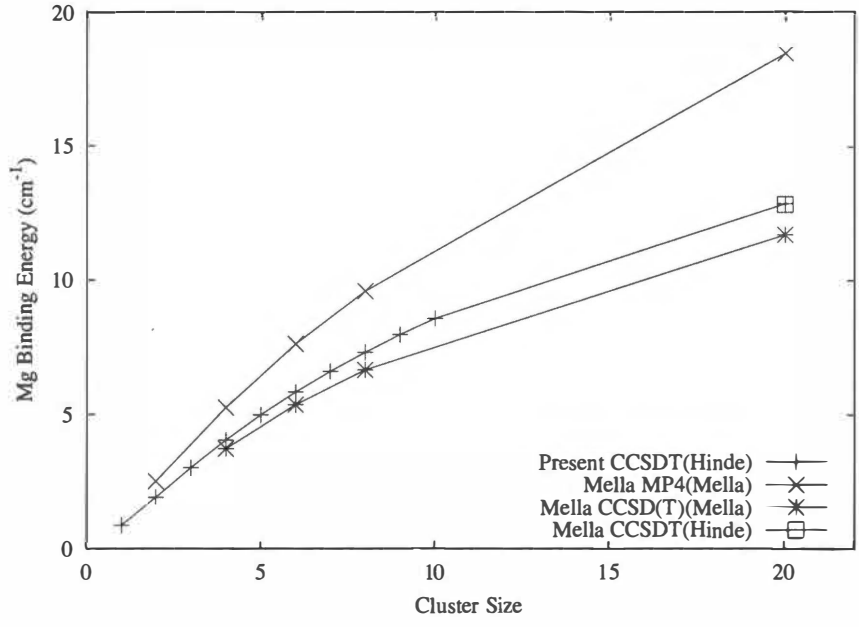


Figure 6.5: Magnesium binding energies (cm^{-1}) as a function of the cluster size (n).

6.3 Structure

Since we have access to the mixed probability density ($\Psi_0\Psi_T$) within a DMC calculation, it is theoretically possible to compute the expectation value of any observable using DMC methods. Thus, we are not limited to estimates of the energy. In the following sections, we use this power of DMC to compute two and three-atom distribution functions in an effort to elucidate the structure of these doped clusters. A commonly used method of generating these distributions is to use second order estimates incorporating both VMC and DMC information to estimate observables of the pure probability density Ψ_0^2. However, the accuracy of this method relies in part on the quality of the trial wavefunction. Here, we use descendant weighting techniques to transform the mixed density into the pure density, giving us direct access to the exact distribution functions.

6.3.1 Radial pair distribution functions

We begin our investigation of the structure of calcium and magnesium doped clusters by examining the radial pair distribution functions of the He-He, He-Ca, and He-Mg interactions. These distribution functions are plotted in figures 6.6, 6.7, 6.8 and 6.9 for the cluster sizes $n = 2$, 5, 10, and 20, respectively. As n increases, the distributions are largely unchanged up to $n = 5$, though perhaps slightly more compact. As n increases to 10 heliums, we begin to see a buildup of density in the tail for the calcium dopant. This effect is likely correlated with the increased "wings" of the calcium density around pairs of helium atoms as seen in the three-particle distributions discussed

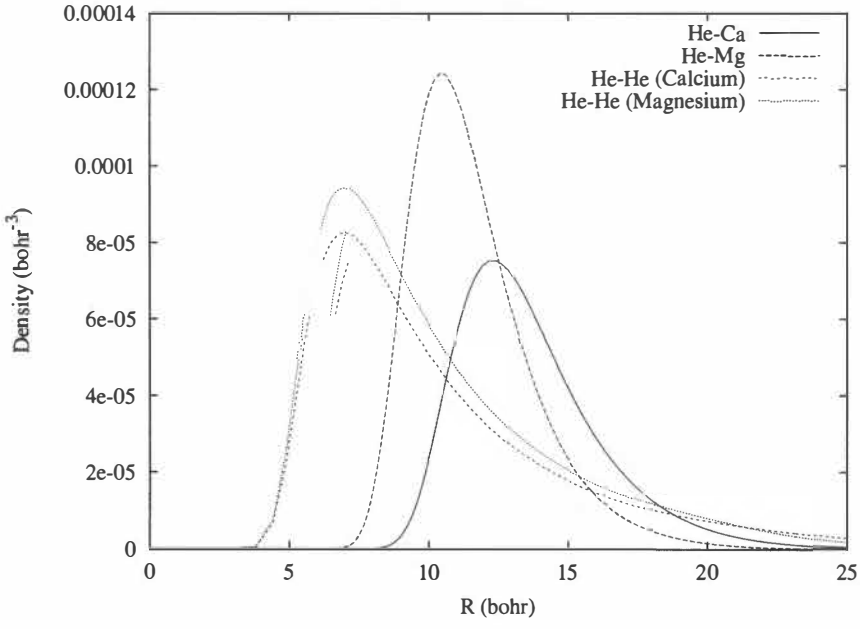


Figure 6.6: He-Mg, He-Ca, and He-He radial distributions in $MgHe_2$ and $CaHe_2$.

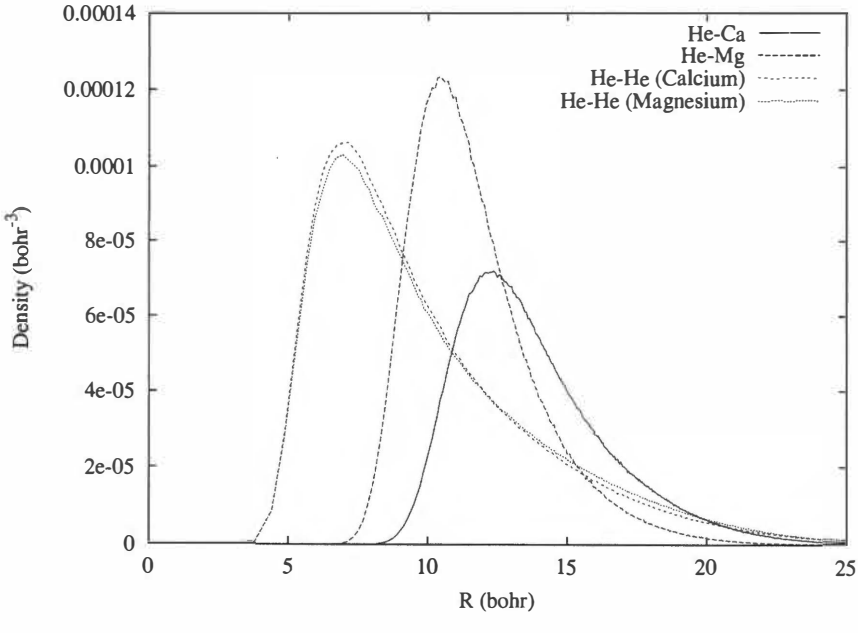


Figure 6.7: He-Mg, He-Ca, and He-He radial distributions in $MgHe_5$ and $CaHe_5$.

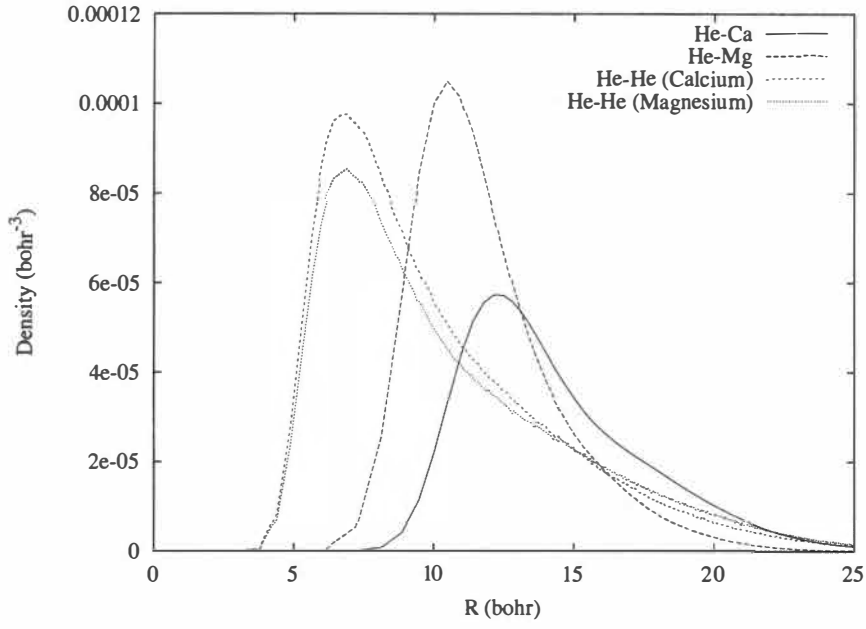


Figure 6.8: He-Mg, He-Ca, and He-He radial distributions in $MgHe_{10}$ and $CaHe_{10}$.

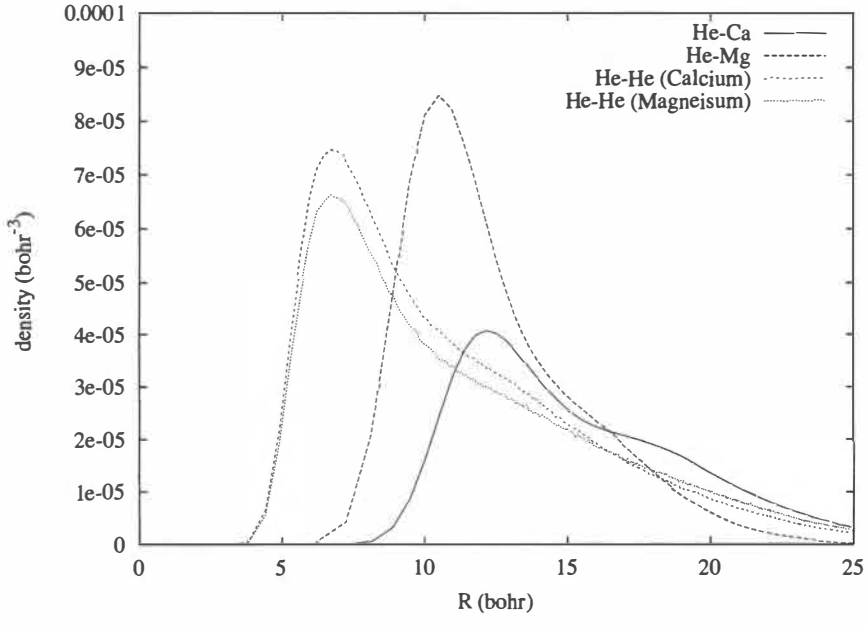


Figure 6.9: He-Mg, He-Ca, and He-He radial distributions in $MgHe_{20}$ and $CaHe_{20}$.

later. Calcium is being expelled further from the helium moiety as the number of heliums increases. The interesting buildup of density in the tail of the He-He density in $MgHe_{10}$ is caused by the arms of the helium density reaching out and around the dopant. In these configurations, there is significant helium density farther away from the central helium moiety. Corroborating this is the fact that the helium pair density in $MgHe_{10}$ has grown near $R = 7$ bohr, indicating a more localized and cluster-like clump of helium. These effects are even more pronounced in the doped He_{20} clusters, with a clear shoulder now emerging in the tail of calcium density.

6.3.2 Distribution functions about the helium center of mass

To directly attack the question of whether a particular dopant resides near the center or on the surface of a helium cluster, it is useful to examine the distribution of dopant distances from the center of mass of the helium moiety. Distances near zero imply that the center of mass of the helium moiety is coincident with the position of the dopant, indicating solvation of the dopant within the center of the cluster. Density profiles centered at larger separations are indicative of off-center or surface solvation sites. Figures C.1-C.10 show the relative distributions of calcium and magnesium distances from the helium center of mass for increasing cluster size n.

For $n = 2$, we see bimodal distributions, characteristic of the dopant between the two heliums in nearly linear arrangements as well as in more triangular configurations. The propensity for linear arrangements is more pronounced for magnesium than for calcium. We also observe that even at $n = 2$, the calcium density distribution already has a noticeably longer tail toward greater distances. However, it is not clear whether this behavior is primarily due to the weaker well or larger equilibrium separation of the Ca-He pair potential.

By $n = 4$, the probability of finding calcium near the center of the cluster is quite small. This contrasts with magnesium, which still shows noticeable density in this region even for $MgHe_8$. For $8 \leq n \leq 10$, the position of the magnesium dopant with respect to the helium center of mass is relatively constant and appears to be roughly 5 bohr closer than for calcium. However, this information alone does not unambiguously imply that the magnesium is solvated deeper within the cluster since the compactness of the helium moiety also needs to be considered. By $n = 20$, the calcium dopant density has shifted even further out while the magnesium has moved slightly inward, with an increased amount of density near the center of the helium moiety.

6.3.3 Dopant density around helium pairs

To better clarify the observed structural features, we examine the density of dopant atoms distributed around pairs of helium atoms (figures D.1-D.20). For each pair of helium atoms, we define a z axis, taking $z = 0$ to be the geometrical midpoint between the two helium atoms. Next, we define the radial coordinate r defined as the perpendicular distance from the z axis and construct a two dimensional histogram in the coordinates z and r, such that the bins have equal volume. Positions of the dopant atoms are then referenced through the cylindrical variables z and r and are binned appropriately for each pair of helium atoms. The resulting two dimensional distribution will

be cylindrically averaged over the azimuthal angle ϕ, and we choose a normalization such that

$$1 = 2\pi \int_{-\infty}^{\infty} dz \int_{0}^{\infty} r dr \rho(r, z). \tag{6.3}$$

Three structural features are evident from these plots. First, for all cluster sizes, the magnesium density is more localized than the corresponding calcium density. Second, we note that the magnesium density between two helium atoms is significantly higher than that of calcium, with this persisting even to the larger cluster sizes. Third, calcium, unlike magnesium, appears to have higher density in"wings" reaching out around both heliums. These wings while small grow with increasing cluster size and are a strong indication of calcium's increased preference for surface binding in the larger clusters. For $n = 20$, we see some dramatic differences between the calcium and magnesium density distributions. For calcium, density in the wings has grown considerably and the peak has broadened substantially. Since the increased density toward the wings arises from linear and nearly linear He-He-Ca arrangements, we must conclude that the heliums are largely aggregated together, separate from the dopant. For $MgHe_{20}$, the trend is reversed, favoring a more compact magnesium density and an increased likelihood of finding the magnesium atom near the center of the helium moiety, compared to $MgHe_{10}$. Based on the evidence of Mella *et al*,[16] this is likely a prelude to the full interior solvation of magnesium observed in larger cluster sizes.

6.3.4 Helium density about the dopant-helium center of mass

The most informative plots of density distributions we present describe the helium density with respect to the geometry of the entire doped cluster. Like the two dimensional density distributions considered in the previous section, we again use a cylindrical coordinate system to characterize helium positions with respect to a z axis taken through the center of mass of the helium atoms and the dopant. The point $z = 0$ is taken to be the geometrical midpoint of the distance between the helium mass center and the dopant, with the dopant atom located on the positive z axis. For convenient comparisons between cluster sizes, we have chosen to normalize the helium density as a "reduced density" according to

$$1 = 2\pi \int_{-\infty}^{\infty} dz \int_{0}^{\infty} r dr \rho(r, z), \tag{6.4}$$

though we could just as easily have required that the density integrate to the number of helium atoms in the cluster.

These density distributions are exhibited in figures E.1-E.20. For the smallest cluster sizes, we see considerable localization of the helium density for both calcium and magnesium doped systems. Since helium atoms wish to establish as many neighboring helium-helium interactions as possible while maintaining a favorable interaction with the dopant, this is not unexpected. By $3 \leq n \leq 5$, we observe considerable deformation of the helium density around the dopant. It also appears that the presence of magnesium introduces a stronger deformation by thinning out the helium density

and wrapping it further around the magnesium atom. This noticeable degree of "cupping" indicates that the magnesium-helium interaction is nearly competitive with the stabilizing helium-helium interactions. In the regime $6 \leq n \leq 10$, this behavior leads to some discernible structural differences. In magnesium doped clusters, the helium density continues to wrap around the dopant while in the calcium case, additional density piles up behind the existing density opposite the dopant. Similar to a second solvation shell, this "cap" or "shelf" of extra density forms without the dopant being fully solvated. This build up of helium density establishes a more cluster-like helium moiety with the calcium further expelled from the interior. In contrast to magnesium, this indicates that the calcium-helium interaction is not as competitive as the more preferred helium-helium interactions.

For the largest cluster sizes studied here ($n = 20$), there are indications that the helium density is beginning to reach around and behind the magnesium dopant. If we assume that magnesium is solvated in the interior of larger clusters ($n > 30$) as Mella *et al* [16] claim, then dopant solvation in the $MgHe_{20}$ system should be controlled by a very delicate energetic balance. The special crossover behavior from exterior to interior solvation for a dopant appears to be extremely sensitive to the employed pair potentials and has only been observed with the CCSDT potential of Hinde. Results for MP4 and CCSD(T) potentials [16] predict either full or no interior solvation of magnesium, respectively, for a wide range of cluster sizes. While subject to experimental verification, the existence and cluster size dependence ofthis unique solvation crossover behavior would provide valuable insight into the finite size effects present in quantum clusters.

6.3.5 Magnesium solvation dependence on the interaction strength

As noted above, the solvation structure of magnesium in $MgHe_{20}$ should be uniquely senstive to external perturbation. To test this hypothesis, we artifically perturb the Mg-He interaction by systematically increasing the well depth of the CCSDT(Hinde) potential. This is accomplished by scaling the potential by a multiplicative factor. We find that only a 3-4% increase in the well depth is sufficient for the interior solvation of magnesium. This is evidenced in figure 6.10 where we have plotted the dopant-He(MC) density distributions for several different well depths. The dramatic shift in solvation behavior with respect to small perturbations in the potential is convincing proof of the delicate energy balance present in $MgHe_{20}$. With the considerable variation present in currently available Mg-He potentials, a precise experimental determination of the minimum cluster size where this crossover effect appears would go a long way toward identifying the appropriate well depth in the Mg-He interaction and aid the construction of better theoretical models.

6.4 Summary

From the relative energetics, it is clear that calcium is less strongly bound to helium clusters than magnesium. This fact could have been qualitatively predicted from the deeper well of the magnesium-helium PES compared to that of calcium. While it is true that weaker binding is associated with surface attached dopants, a solid case for the interior or surface solvation of these

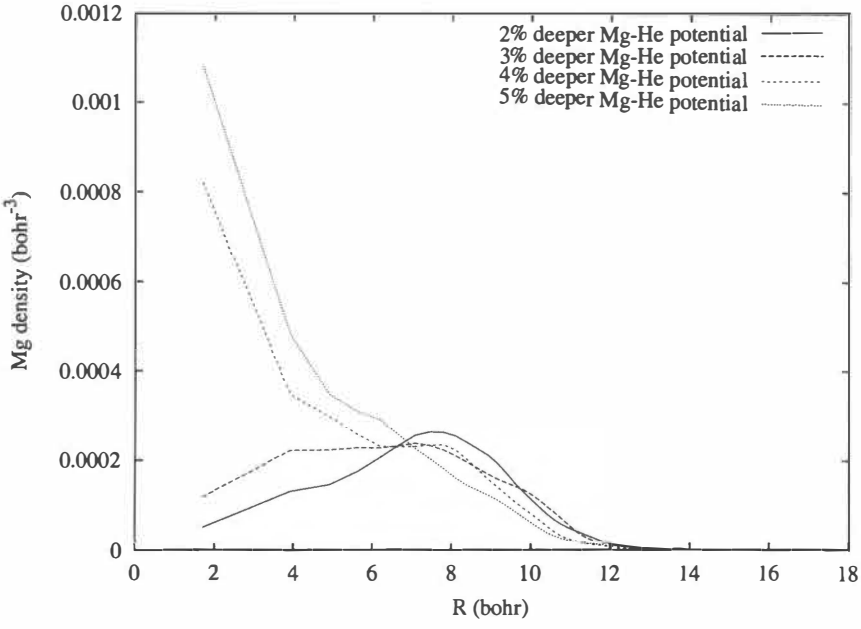


Figure 6.10: Mg-He(MC) distributions in $MgHe_{20}$ with increasing well depth.

dopants cannot be made on the basis of binding energies alone. Large zero point energies make a considerable difference, since the deeper He-He potential well would classically maximize the number of He-He interactions at the expense of the He-dopant interactions. Consequently, it is thought that the interior solvation of magnesium in these clusters is a purely quantum effect.

The two and three-particle distribution functions that we have computed in this chapter establish a more compelling case that these dopants reside in deep wells or pockets near the surface of the cluster for $n \leq 20$. The case for magnesium appears less clear-cut than that for calcium since there are indications that magnesium may become more fully solvated as the cluster sizes are increased. However, in the present study it is clear that magnesium atoms can and do make excursions to the center of the droplet for all of the cluster sizes studied here. There is no indication of similar behavior for calcium dopants.

Chapter 7

Conclusions and Summary

7.1 Conclusions

In chapter three, we developed a framework for expressing the statistical error as a function of population size, importance sampling, and dimensionality. For arbitrary operators in high dimensional systems, we find exponential scaling of the statistical variance for all but perfect importance sampling. This poses serious restrictions on the quality of trial wavefunctions for many-body systems with many degrees of freedom. We have also disambiguated the roles of the statistical and quantum mechanical variances for a given operator and the sharing of zero-variance properties.

In chapter four, we apply our theory to an investigation of the k-dimensional isotropic harmonic oscillator as a first approximation of high dimensional cluster systems. We find that the statistical distribution of operator values are inherently non-Gaussian for small population sizes and that, contrary to intuition, importance sampling does not make these distributions more Gaussian. In fact, importance sampling often introduces significant skewness even into initially symmetric distributions. From this we conclude that large populations are absolutely necessary to obtain accurate estimates of properties, even in importance sampled calculations. In this chapter, we also highlight the dimensional dependence of certain common operators demonstrating explicit zero-variance and non-zero-variance behavior for these estimators.

In chapter five, we develop a new trial function *ansatz* for the investigation of pure and metal doped helium clusters. We optimize these wavefunctions using novel energy optimization techniques to produce high quality wavefunctions amenable to the present study of doped helium clusters. We also discuss the importance of Gaussian statistics and central limit theorem behavior in the design of robust optimization procedures.

In chapter six, we perform diffusion Monte Carlo calculations on pure and doped helium clusters to determine the total energies, helium evaporation energies, and dopant binding energies in these systems. The results we present for the calcium doped clusters are the first ever calculations for this system. We also employ two and three-particle density distribution functions to elucidate differences in the solvation behavior for calcium and magnesium dopants. The results are in agree-

ment with recent theoretical work as well as current experimental findings, predicting that calcium will bind less strongly than magnesium and be solvated near the surface of the cluster in a deep well or pocket, while magnesium prefers a much deeper or possibly an interior location.

7.2 Future Work

Future work will continue to focus on both the theoretical development of QMC methods and their successful and accurate application to physically interesting problems. Concerning the theory of QMC, the computation of observables with reduced statistical error and bias continues to be problematic. Future investigations in this area would include an assessment of the statistical error and bias inherent in descendant weighting techniques in small sampling populations and the development of recently introduced techniques for constructing zero-variance, low bias estimators for properties. We would also like to see the continued development of stable and robust optimization methods, specifically for energy optimization.

Regarding the practical application of QMC techniques, we foresee investigations of larger helium droplets to verify the recently observed crossover behavior in the solvation of magnesium dopants. Such investigations would be valuable for characterizing the role of finite size effects and useful for improving the description of the λ solvation model in these systems. We would also like to develop additional tools for the calculation and prediction of more physically accessible quantities, such as frequency shifts and spectroscopic profiles.

LIST OF REFERENCES

List of References

[1] J. A. Northby, J. Chem. Phys. **115**, 10065 (2001).

[2] M. A. McMahon, R. N. Barnett, and K. B. Whaley, J. Chem. Phys. **99**, 8816 (1993).

[3] J. Gspann, Phys B: Condens. Matt. **98**, 405 (1995).

[4] A. Scheidemann, J. P. Toennies, and J. A. Northby, Phys. Rev. Lett. **64**, 1899 (1990).

[5] D. Blume, M. lewerenz, F. Huisken, and M. Kaloudis, J. Chem. Phys. **105**, 8666 (1996).

[6] S. Moroni, N. Blinov, and P. Roy, J. Chem. Phys. **121**, 3577 (2004).

[7] R. N. Barnett and K. B. Whaley, J. Chem. Phys. **99**, 9730 (1993).

[8] F. Stienkemeier, J. Higgins, C. Callegari, S. I. Kanorsky, W. E. Ernst, and G. Scoles, Z. Phys. D. **38**, 253 (1996).

[9] F. Ancilotto, E. Cheng, M. W. Cole, and F. Toigo, Z. Phys. B. **98**, 823 (1995).

[10] F. Stienkemeier, F. Meier, and H. O. Lutz, J. Chem. Phys. **107**, 10816 (1997).

[11] Y. Takahashi, K. Sano, T. Kinoshita, and T. Yabuzaki, Phys. Rev. Lett. **71**, 1035 (1993).

[12] J. Reho, U. Merker, M. R. Radcliff, K. K. Lehmann, and G. Scoles, J. Chem. Phys. **112**, 8409 (2000).

[13] Y. Moriwaki and N. Morita, Eur. Phys. J. D **5**, 53 (1999).

[14] R. J. Hinde, J. Phys. B **36**, 3119 (2003).

[15] H. Partridge, J. R. Stallcop, and E. Levin, J. Chem. Phys. **115**, 6471 (2001).

[16] M. Mella, G. Calderoni, and F. Cargnoni (2005), unpublished (arXiv.org:physics/0502022).

[17] F. Ancilotto, P. B. Lerner, and M. W. Cole, J. Low. Temp. Phys. **101**, 1123 (1995).

[18] V. R. Pandharipande, J. G. Zabloitzky, S. C. Pieper, R. B. Wiringa, and U. Helmbrecht, Phys. Rev. Lett. **50**, 1676 (1983).

[19] M. H. Kalos, M. A. Lee, P. A. Whitlock, and G. V. Chester, Phys. Rev. B **24**, 115 (1981).

[20] R. Melzer and J. G. Zabolitzky, J. Phys A: Math. Gen. **17**, L565 (1984).

[21] S. W. Rick, D. L. Lynch, and J. D. Doll, J. Chem. Phys. **95**, 3506 (1991).

[22] S. A. Chin and E. Krotscheck, Phys. Rev. B **45**, 852 (1992).

[23] M. V. R. Krishna and K. B. Whaley, Phys. Rev. Lett. **64**, 1126 (1990).

[24] R. N. Barnett and K. B. Whaley, Phys. Rev. A **47**, 4082 (1993).

[25] K. B. Whaley, Int. Rev. Phys. Chem. **13**, 41 (1994).

[26] R. N. Barnett and K. B. Whaley, J. Chem. Phys. **99**, 9730 (1993).

[27] M. Lewerenz, J. Chem. Phys. **106**, 4596 (1997).

[28] M. V. Patel, A. Viel, F. Paesani, P. Huang, and K. B. Whaley, J. Chem. Phys. **118**, 5011 (2003).

[29] A. Viel and K. B. Whaley, J. Chem. Phys. **115**, 10186 (2001).

[30] N. Metropolis and S. Ulam, J. Am. Stat. Assoc. **44**, 335 (1949).

[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, M. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[32] G. Buffon, *Histoire naturelle, générale er particulière* (1777).

[33] R. Bellman, *Adaptive Control Processes: A Guided Tour* (Princeton University Press, 1961).

[34] P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to stochastic processes* (Houghton Mifflin, 1987).

[35] D. Bressanini and P. J. Reynolds, in *Monte Carlo Methods in Chemical Physics*, edited by D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar (Wiley, 1999), vol. 105.

[36] S. A. Chin, Phys. Rev. A **42**, 6991 (1990).

[37] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).

[38] A. Mira, Ph.D. thesis, University of Minnesota (1998).

[39] L. Tierney and A. Mira, Statist. Med. **18**, 2507 (1999).

[40] P. J. Green and A. Mira, Biometricka **88**, 1035 (2001).

[41] D. Bressanini, G. Morosi, S. Tarasco, and A. Mira, J. Chem. Phys. **121**, 3446 (2004).

[42] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).

[43] J. B. Anderson, J. Chem. Phys. **73**, 3897 (1980).

[44] A. J. James, Ph.D. thesis, University of London (1995).

[45] R. N. Barnett, P. J. Reynolds, and J. W. A. Lester, J. Comp. Phys. **96**, 258 (1991).

[46] J. Casulleras and J. Boronat, Phys. Rev. B **52**, 3654 (1995).

[47] R. Assaraf, M. Caffarel, and A. Khelif, Phys. Rev. E **61**, 4566 (2000).

[48] I. Bosa and S. M. Rothstein, J. Chem. Phys. **121**, 4486 (2004).

[49] K. S. Liu, M. H. Kalos, and G. V. Chester, Phys. Rev. A **10**, 303 (1974).

[50] P. J. Reynolds, R. N. Barnett, B. L. Hammond, and J. W. A. Lester, J. Stat. Phys. **43**, 1017 (1986).

[51] K. J. Runge and R. J. Runge, Phys. Rev. B **45**, 7229 (1992).

[52] M. Caffarel and P. Claverie, J. Chem. Phys. **88**, 1088 (1988).

[53] R. Assaraf and M. Caffarel, J. Chem. Phys. **119**, 10536 (2003).

[54] A. Sarsa, J. Boronat, and J. Casulleras, J. Chem. Phys. **116**, 5956 (2002).

[55] J. Boronat and J. Casulleras, Phys. Rev. B **59**, 8844 (1999).

[56] N. V. Smirnov, Bulletin Moscow University **2**, 3 (1939).

[57] T. W. Anderson and D. A. Darling, J. Am. Stat. Assoc. **49**, 765 (1954).

[58] W. G. Cochran, Ann. Math. Stat. **23**, 315 (1952).

[59] I. H. Sloan and H. Wozniakowski, Tech. Rep. AMR03/11, University of New South Wales (2003).

[60] J. Rust, J. F. Traub, and H. Wozniakowski, Econometrica **70**, 285 (2002).

[61] F. Hickernell and H. Wozniakowski, Adv. Comput. Math. **12**, 25 (2000).

[62] J. Rust, Econometrica **65**, 487 (1997).

[63] R. Assaraf and M. Caffarel, J. Chem. Phys. **113**, 4028 (2003).

[64] R. Assaraf and M. Caffarel, Phys. Rev. Lett. **83**, 4682 (1999).

[65] R. Guardiola, M. Portesi, and J. Navarro, Phys. Rev. B **60**, 6288 (1999).

[66] W. L. McMillan, Phys. Rev. **138**, A442 (1965).

[67] M. A. McMahon and K. B. Whaley, J. Chem. Phys. **182**, 119 (1994).

[68] S. Chiesa, M. Mella, G. Morosi, and D. Bressanini, J. Chem. Phys. **119**, 5601 (2003).

[69] R. A. Aziz, F. R. W. McCourt, and C. C. K. Wong, Mol. Phys. **61**, 1487 (1987).

[70] M. Casalegno, Ph.D. thesis, Milano (2001).

[71] D. Bressanini, G. Morosi, and M. Mella, J. Chem. Phys. **116**, 5345 (2002).

[72] M. Snajdr and S. M. Rothstein, J. Chem. Phys. **112**, 4935 (2000).

[73] X. Lin, H. K. Zhang, and A. M. Rappe, J. Chem. Phys. **112**, 2650 (2000).

[74] F. J. Galvez, E. Buendia, and A. Sarsa, J. Chem. Phys. **115**, 1166 (2001).

[75] P. R. C. Kent, Ph.D. thesis, University of Cambridge (1999).

[76] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).

[77] J. Nocedal and S. Wright, *Numerical Optimization* (Springer Verlag, 1999).

[78] A. R. Janzen and R. A. Aziz, J. Chem. Phys. **107**, 914 (1997).

[79] R. A. Aziz, A. R. Janzen, and M. R. Moldover, Phys. Rev. Lett. **74**, 1586 (1995).

[80] K. T. Tang, J. P. Toennies, and C. L. Yiu, Phys. Rev. Lett. **74**, 1546 (1995).

[81] T. van Mourik and T. H. D. Jr., J. Chem. Phys. **111**, 9248 (1999).

[82] J. B. Anderson, J. Chem. Phys. **115**, 4546 (2001).

[83] D. M. Ceperley and H. J. Partridge, J. Chem. Phys. **84**, 820 (1986).

[84] T. van Mourik and J. H. van Lenthe, J. Chem. Phys. **103**, 9626 (1995).

[85] K. T. Tang and J. P. Toennies, J. Chem. Phys. **80**, 3726 (1984).

[86] T. Korona, H. L. Williams, R. Bukowski, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **106**, 1 (1997).

[87] D. Bishop and J. Pipin, Int. J. Quant. Chem. **45**, 349 (1993).

[88] J. Mitroy and M. W. J. Bromley, Phys. Rev. A **68**, 62710 (2003).

[89] J. M. Standard and P. R. Certain, J. Chem. Phys. **83**, 3002 (1985).

[90] C. C. Lovallo and M. Klobukowski, Chem. Phys. Lett. **373**, 439 (2003).

[91] D. J. Funk, W. H. Brekenridge, J. Simons, and G. Chalasinski, J. Chem. Phys. **91**, 1114 (1989).

[92] M. G. Kendall, *The Advanced Theory of Statistics*, vol. 1 (Charles Griffen, 1945), 2nd ed.

APPENDICES

Appendix A

Proofs

A.1 Proof of Equation 3.37 for $\langle\Omega\rangle_k$

Beginning with an isotropic product distribution (equation 3.2), the projector definition

$$P_n(\mathbf{x}) = \prod_{i=1}^{k} P_n(x_i),$$

and a generic k-dimensional coordinate operator as defined in equation 3.36, we can write the expectation value of that operator (equation 3.7) as

$$\langle\Omega\rangle_k = \int \hat{P}_n(\mathbf{x})\hat{O}\pi(\mathbf{x})d\mathbf{x} = \int \left(\prod_{j=1}^{k} P_n(x_j)\right)\left(\frac{1}{k}\sum_{i=1}^{k}\hat{O}_i\right)\left(\prod_{\ell=1}^{k}\pi(x_\ell)\right)d\mathbf{x}.$$

Now, we move the summation to the outside and remove all terms not involving x_i from the innermost integral, giving

$$\langle\Omega\rangle_k = \frac{1}{k}\sum_{i=1}^{k}\int \left(\frac{1}{P_n(x_i)}\prod_{j=1}^{k} P_n(x_j)\right)\left(\frac{1}{\pi(x_i)}\prod_{\ell=1}^{k}\pi(x_\ell)\right)\frac{d\mathbf{x}}{dx_i}\underbrace{\int P_n(x_i)\,\hat{O}_i\,\pi(x_i)dx_i}_{\mathcal{I}_{1,1}(x_i)}.$$

Next, we rewrite the integral of products as the product of integrals

$$\langle\Omega\rangle_k = \frac{1}{k}\sum_{i=1}^{k}\mathcal{I}_{1,1}(x_i)\prod_{j=1,j\neq i}^{k}\underbrace{\int P_n(x_j)\pi(x_j)dx_j}_{\mathcal{I}_{1,0}(x_j)}.$$

Since we have chosen to make $\pi(\mathbf{x})$ isotropic, all of the one dimensional $\mathcal{I}_{1,0}$ integrals are identical and we can express them in powers of k as

$$\langle \Omega \rangle_k = \frac{1}{k} \sum_{i=1}^{k} \mathcal{I}_{1,1}(x_i) \mathcal{I}_{1,0}^{k-1}.$$

Likewise, the integrals $\mathcal{I}_{1,1}$ are all identical, and the expression reduces to

$$\langle \Omega \rangle_k = \mathcal{I}_{1,1} \mathcal{I}_{1,0}^{k-1}$$

A.2 Proof of Equation 3.38 for $\langle V_{N,M}^2 \rangle_k$

Since $\hat{\mathcal{O}}$ is a coordinate operator, commutativity holds and the product $(P_n \hat{\mathcal{O}})^2$ may be rewritten as $P_n^2 \hat{\mathcal{O}}^2$. Using the approach of proof A.1, we may write

$$\hat{\mathcal{O}}^2 = \left(\frac{1}{k} \sum_{i=1}^{k} \hat{O}_i \right)^2 = \frac{1}{k^2} \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{O}_i \hat{O}_j.$$

We can partition $\hat{\mathcal{O}}^2$ into a sum of two operators, $\hat{A}$ and $\hat{B}$, according to

$$\hat{\mathcal{O}}^2 = \underbrace{\frac{2}{k^2} \sum_{i=1}^{k} \sum_{j>i}^{k} \hat{O}_i \hat{O}_j}_{\hat{A}} + \underbrace{\frac{1}{k^2} \sum_{i=1}^{k} \hat{O}_i^2}_{\hat{B}}.$$

The expectation value of $P_n^2 \hat{A}$ is given by

$$\langle P_n^2 A \rangle_k = \int \left(\prod_{m=1}^{k} P_n(x_m) \right)^2 \left(\frac{2}{k^2} \sum_{i=1}^{k} \sum_{j>i}^{k} \hat{O}_i \hat{O}_j \right) \left(\prod_{\ell=1}^{k} \pi(x_\ell) \right) d\mathbf{x}.$$

Now, we pull the summations outside the integration and take all terms not involving i or j out of the innermost integral, leaving

$$\begin{aligned} \langle P_n^2 A \rangle_k \quad = \quad & \frac{2}{k^2} \sum_{i=1}^{k} \sum_{j>i}^{k} \int \left(\frac{1}{P_n(x_i) P_n(x_j)} \prod_{m=1}^{k} P_n(x_m) \right)^2 \left(\frac{1}{\pi(x_i)\pi(x_j)} \prod_{\ell=1}^{k} \pi(x_\ell) \right) \frac{d\mathbf{x}}{dx_i dx_j} \\ & \int P_n^2(x_i)\, P_n^2(x_j)\, \hat{O}_i \hat{O}_j \pi(x_i) \pi(x_j) dx_i\, dx_j. \end{aligned}$$

Since the summation assures us that $j \neq i$, we can rewrite the innermost integral as a product of two integrals

$$\langle P_n^2 A\rangle_k = \frac{2}{k^2}\sum_{i=1}^{k}\sum_{j>i}^{k}\int\left(\frac{1}{P_n(x_i)P_n(x_j)}\prod_{m=1}^{k}P_n(x_m)\right)^2\left(\frac{1}{\pi(x_i)\pi(x_j)}\prod_{\ell=1}^{k}\pi(x_\ell)\right)\frac{d\mathbf{x}}{dx_i dx_j}$$
$$\underbrace{\int P_n^2(x_i)\hat{O}_i\pi(x_i)dx_i}_{\mathcal{I}_{2,1}(x_i)}\underbrace{\int P_n^2(x_j)\hat{O}_j\pi(x_j)dx_j}_{\mathcal{I}_{2,1}(x_j)}.$$

Rewriting the integral of products as a product of integrals, we have

$$\langle P_n^2 A\rangle_k = \frac{2}{k^2}\sum_{i=1}^{k}\sum_{j>i}^{k}\mathcal{I}_{2,1}(x_i)\mathcal{I}_{2,1}(x_j)\prod_{\ell=1,l\neq i,j}^{k}\underbrace{\int P_n^2(x_\ell)\,\pi(x_\ell)dx_\ell}_{\mathcal{I}_{2,0}(x_\ell)}.$$

Finally, recognizing that these one dimensional integrals are identical (because $\pi(\mathbf{x})$ is isotropic), we can express them in terms of k as

$$\langle P_n^2 A\rangle_k = \frac{2}{k^2}\frac{k(k-1)}{2}\mathcal{I}_{2,1}^2\mathcal{I}_{2,0}^{k-2} = \frac{k-1}{k}\mathcal{I}_{2,1}^2\mathcal{I}_{2,0}^{k-2}.$$

For the operator $P_n^2\hat{B}$, we have the expected value

$$\langle P_n^2 B\rangle_k = \frac{1}{k^2}\sum_{i=1}^{k}\int\left(\prod_{j=1}^{k}P_n(x_j)\right)^2\hat{O}_i^2\left(\prod_{\ell=1}^{k}\pi(x_\ell)\right)d\mathbf{x}$$

which we can simplify by pulling all terms not involving i out of the innermost integral to give

$$\langle P_n^2 B\rangle_k = \frac{1}{k^2}\sum_{i=1}^{k}\int\left(\frac{1}{P_n(x_i)}\prod_{j=1}^{k}P_n(x_j)\right)^2\left(\frac{1}{\pi(x_i)}\prod_{\ell=1}^{k}\pi(x_\ell)\right)\frac{d\mathbf{x}}{dx_i}\underbrace{\int P_n^2(x_i)\hat{O}_i^2\pi(x_i)d\mathbf{x}}_{\mathcal{I}_{2,2}(x_i)}.$$

Now, we again rewrite the integral of products as a product of integrals

$$\langle P_n^2 B\rangle_k = \frac{1}{k^2}\sum_{i=1}^{k}\mathcal{I}_{2,2}(x_i)\prod_{\ell=1,\ell\neq i}^{k}\underbrace{\int P_n^2(x_\ell)\pi(x_\ell)dx_\ell}_{\mathcal{I}_{2,0}(x_\ell)}$$

and express them in powers of k to obtain

$$\langle P_n^2 B\rangle_k = \frac{1}{k}\mathcal{I}_{2,2}\mathcal{I}_{2,0}^{k-1}.$$

The variance of $\hat{\Omega} = \hat{P}_n \hat{O}$ is given by

$$\langle V_{N,M}^2 \rangle_k = \langle \Omega^2 \rangle_k - \langle \Omega \rangle_k^2 = \langle P_n^2 B \rangle_k + \langle P_n^2 A \rangle_k - \langle \Omega \rangle_k^2.$$

Substituting in, we are left with

$$\langle V_{N,M}^2 \rangle_k = \frac{1}{k} \mathcal{I}_{2,2} \mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k} \mathcal{I}_{2,1}^2 \mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2 \mathcal{I}_{1,0}^{2k-2}.$$

From statistics, we know that the variance of a random (independent) sample of size N is just the variance for a single sample divided by the sample size. Making use of this fact, we have

$$\langle V_{N,M}^2 \rangle_k = \frac{1}{N} \left\{ \frac{1}{k} \mathcal{I}_{2,2} \mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k} \mathcal{I}_{2,1}^2 \mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2 \mathcal{I}_{1,0}^{2k-2} \right\}.$$

A.3 Proof of Equation 3.38 for $\langle \bar{S}_{N,M}^2 \rangle_k$

In an effort to avoid needless duplication of work and tedious algebra here and in future proofs, we draw upon known results from statistical theory regarding the moments of a distribution. In one dimension, a central moment μ_b is defined as the expectation value

$$\mu_b = \langle (\hat{\Omega} - \langle \Omega \rangle)^b \rangle$$

taken about the population mean $\langle \Omega \rangle$, where $\hat{\Omega} = P_n \hat{O}$. These central moments are related to raw moments about zero via the relations

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= \langle \Omega^2 \rangle - \langle \Omega \rangle^2 \\
\mu_4 &= \langle \Omega^3 \rangle - 3\langle \Omega^2 \rangle \langle \Omega \rangle + 2\langle \Omega \rangle^3 \\
\mu_4 &= \langle \Omega^4 \rangle - 4\langle \Omega^3 \rangle \langle \Omega \rangle + 6\langle \Omega^2 \rangle \langle \Omega \rangle^2 - 3\langle \Omega \rangle^4.
\end{aligned}$$

In k dimensions, the central moments are expressed as

$$\begin{aligned}
\mu_1 &= 0 \\
\mu_2 &= \langle \Omega^2 \rangle_k - \langle \Omega \rangle_k^2 \\
\mu_4 &= \langle \Omega^3 \rangle_k - 3\langle \Omega^2 \rangle_k \langle \Omega \rangle_k + 2\langle \Omega \rangle_k^3 \\
\mu_4 &= \langle \Omega^4 \rangle_k - 4\langle \Omega^3 \rangle_k \langle \Omega \rangle_k + 6\langle \Omega^2 \rangle_k \langle \Omega \rangle_k^2 - 3\langle \Omega \rangle_k^4.
\end{aligned}$$

where

$$\langle \Omega \rangle_k = \mathcal{I}_{1,1} \mathcal{I}_{1,0}^{k-1},$$

$$\langle \Omega^2 \rangle_k = \frac{1}{k} \mathcal{I}_{2,2} \mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k} \mathcal{I}_{2,1}^2 \mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2 \mathcal{I}_{1,0}^{2k-2},$$

$$\langle \Omega^3 \rangle_k = \frac{1}{k^2} \mathcal{I}_{3,3} \mathcal{I}_{3,0}^{k-1} + \frac{k-1}{k^2} \mathcal{I}_{3,2} \mathcal{I}_{3,1} \mathcal{I}_{3,0}^{k-2} + \frac{k-1}{k} \mathcal{I}_{3,1}^3 \mathcal{I}_{3,0}^{k-3},$$

and

$$\langle \Omega^4 \rangle_k = \frac{1}{k^3} \mathcal{I}_{4,4} \mathcal{I}_{4,0}^{k-1} + \frac{k-1}{k^3} \left(\mathcal{I}_{4,3} \mathcal{I}_{4,1} + \mathcal{I}_{4,2}^2 \right) \mathcal{I}_{4,0}^{k-2} + \frac{k-1}{k^2} \mathcal{I}_{4,2} \mathcal{I}_{4,1}^2 \mathcal{I}_{4,0}^{k-3} + \frac{k-1}{k} \mathcal{I}_{4,1}^4 \mathcal{I}_{4,0}^{k-4}$$

From statistics,[92] we know that the sample variance operator

$$\hat{S}_{N,M}^2 = \frac{1}{M-1} \sum_{i=1}^{M} \left(\bar{\Omega}_m - \langle \bar{\Omega} \rangle \right)^2$$

has the expectation value

$$\langle \bar{S}_{N,M}^2 \rangle = \frac{M}{M-1} \mu_2$$

Again, taking into account the variance of a random sample of size N, we have

$$\langle \bar{S}_{N,M}^2 \rangle = \frac{1}{N} \left(\frac{M}{M-1} \right) \left\{ \frac{1}{k} \mathcal{I}_{2,2} \mathcal{I}_{2,0}^{k-1} + \frac{k-1}{k} \mathcal{I}_{2,1}^2 \mathcal{I}_{2,0}^{k-2} - \mathcal{I}_{1,1}^2 \mathcal{I}_{1,0}^{2k-2} \right\}$$

or, more simply

$$\langle \bar{S}_{N,M}^2 \rangle_k = \frac{M}{M-1} \langle V_{N,M}^2 \rangle_k.$$

Thus, we see that as $M \to \infty$, $\langle \bar{S}_{N,M}^2 \rangle_k$ is an unbiased estimator for $\langle V_{N,M}^2 \rangle_k$.

A.4 Proof of Equation 3.39 for $\langle \mathbf{V}_L^2 \rangle_k$

Here, we are looking for the expectation value $\langle \mathbf{V}_L^2 \rangle_k$, which from statistics is related to $\mu_4 - \mu_2^2$. However, for sampling from a population of size N, μ_4 is no longer described by the simple central moment

$$\mu_4 = \langle (\hat{\Omega} - \langle \Omega \rangle)^4 \rangle$$

but by

$$\mu_4^N = \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} \hat{\Omega}_i - \langle \Omega \rangle \right)^4 \right\rangle.$$

This new moment may be rewritten as

$$\mu_4^N = \frac{1}{N^4} \left\langle \left(\sum_{i=1}^{N} \hat{\Omega}_i - \langle \Omega \rangle \right)^4 \right\rangle.$$

and expanded as simple moments to give

$$\mu_4^N = \frac{1}{N^3} \Big(\mu_4 + 3(N-1) \mu_2^2 \Big).$$

Likewise, μ_2 must be replaced by μ_2/N. Making the substitutions for μ_4 and μ_2,, we obtain

$$\langle \mathbf{V}_L^2 \rangle_k = \frac{\mu_4}{N^3} + \frac{3(N-1)\mu_2^2}{N^3} - \frac{\mu_2^2}{N^2}.$$

Rearranging, and taking into account M snapshots, we are left with

$$\langle \mathbf{V}_L^2 \rangle_k = \frac{1}{MN^3} \left[\mu_4 + (2N-3)\mu_2^2 \right].$$

A.5 Proof of Equation 3.45 for $\langle \bar{\mathbf{S}}_L^2 \rangle_k$

Likewise, the variance of the sample variance operator is known[92] to be given by

$$\left(\frac{L}{L-1} \right) \left[\frac{1}{M} \mu_4 - \frac{(M-3)}{M(M-1)} \mu_2^2 \right].$$

Adjusting these moments for a population of size N leads to

$$\langle \bar{\mathbf{S}}_L^2 \rangle_k = \left(\frac{L}{L-1} \right) \left[\frac{\mu_4}{N^3} + \frac{3(N-1)\mu_2^2}{N^3} - \frac{(M-3)\mu_2^2}{(M-1)N^2} \right].$$

Rearranging, we obtain

$$\langle \bar{\mathbf{S}}_L^2 \rangle_k = \left(\frac{L}{L-1} \right) \frac{1}{MN^3} \left[\mu_4 + \left(\frac{2NM}{M-1} - 3 \right) \mu_2^2 \right].$$

A.6 Proof of Equation 4.5

We begin with the differential relationship

$$C_{0,k}^{\alpha}(c)dc = \pi(\mathbf{x}) S_k d\mathbf{x}$$

which relates a change in the coefficient distribution to a change in the stationary distribution. For the harmonic oscillator, $\pi(\mathbf{x})$ is a k-dimensional isotropic (spherical) Gaussian and may be rewritten in terms of the hyper-radius $r = \sqrt{\mathbf{x}^2}$. Thus we take $\pi(\mathbf{x})$ to be

$$\pi(r) = \sqrt{\left(\frac{\alpha+1}{2\pi} \right)^k} e^{-(\alpha+1)r^2/2}$$

and the ground state projector to be

$$P_0(r) = \sqrt{\left(\frac{2}{\alpha+1} \right)^k} e^{-(1-\alpha)r^2/2}.$$

Since $c = P_0(r)$, the differential dc is given by

$$dc = \left(\frac{2}{\alpha+1}\right)^k \pi^{k/2}(\alpha-1)e^{2\alpha r^2/2}\pi(r)rdr.$$

The factor $S_k dr$ is the volume differential in k dimensions. By itself, S_k is the surface area of a k-dimensional hypersphere of radius r, and is equal to

$$S_k = \frac{2\pi^{k/2}r^{k-1}}{\Gamma(k/2)}.$$

Inserting these pieces into the differential relationship, we obtain

$$C^{\alpha}_{0,k}(c) = \frac{(\alpha+1)^k}{2^{k-1}(\alpha-1)\Gamma(k/2)} r^{(k-2)} e^{-(1-\alpha)r^2/2}.$$

We can also take the inverse of $P_0(r)$ to find r

$$r = \sqrt{\frac{1}{\alpha-1}\ln\left\{\left(\frac{\alpha+1}{2}\right)^k c^2\right\}}$$

which can be inserted into the above expression to yield

$$C^{\alpha}_{0,k}(c) = \frac{(\alpha+1)^k w^{\alpha/(\alpha-1)}}{2^{k-1}(\alpha-1)\Gamma(k/2)} \left(\frac{\ln w}{1-\alpha}\right)^{(k-2)/2}$$

where

$$w = c^2\left(\frac{\alpha+1}{2}\right)^k.$$

Appendix B

Optimized Wavefunction Parameters

Table B.1: Optimized trial function parameters for the HFD-B(87) potential in pure helium clusters.

	He_2	He_3	He_4	He_5	He_6	He_7	He_8	He_9	He_{10}	He_{20}
a_0	-0.965634	-0.640391	-0.25074	-0.218108	-0.13005	-0.131985	-0.115305	-0.109629	-0.109712	-0.11555
a_1	-0.00677339	-0.030967	-0.0485237	-0.0444566	-0.043448	-0.0427382	-0.0411308	-0.0409077	-0.0403819	-0.0423306
b	1.0	1.0	0.997499	1.0	1.00436	0.971598	0.962946	0.935732	0.915712	0.714182
α	0.854001	0.813625	0.939501	0.954598	0.96338	0.967472	0.973322	0.972676	0.970175	0.964606
β	3.09373e-06	-0.000140742	0.000288993	0.000292133	0.000246954	0.000250692	0.000238513	0.00026292	0.000264523	0.000287731
t_0	-248.154	-212.855	-238.126	-242.567	-251.692	-251.051	-251.594	-254.217	-255.134	-255.222
t_1	60.1024	50.9051	41.7395	42.9769	41.3374	41.3539	40.9441	41.2796	41.7483	42.7146
t_2	-5.35092	-4.15925	-2.904	-3.46819	-2.77137	-2.85516	-2.85657	-2.88876	-2.90788	-3.2075

Table B.2: Optimized trial function parameters for the HFD-B3-FCI1(95) potential in pure helium clusters.

	He_2	He_3	He_4	He_5	He_6	He_7	He_8	He_9	He_{10}	He_{20}
a_0	-0.962474	-0.651056	-0.24632	-0.214804	-0.148658	-0.129371	-0.115303	-0.109067	-0.109723	-0.11521
a_1	-0.0067047	-0.0303858	-0.0482949	-0.0453708	-0.0435068	-0.0435899	-0.0411461	-0.0408066	-0.0403911	-0.0422872
b	1.0	1.0	1.00059	0.995074	0.996861	0.970135	0.961625	0.938943	0.916432	0.716665
α	0.859759	0.799577	0.938567	0.949761	0.972615	0.968548	0.970056	0.971636	0.969139	0.962906
β	1.51913e-05	-0.000145126	0.000264833	0.000296366	0.000229821	0.000250772	0.000237847	0.000262998	0.000264526	0.000287729
t_0	-250.578	-212.622	-239.67	-244.937	-249.719	-253.614	-251.859	-255.227	-255.238	-255.238
t_1	59.6131	51.5242	41.1804	43.2071	41.2041	41.4153	41.1033	41.3111	41.695	42.7878
t_2	-5.22411	-4.21734	-2.76692	-3.19162	-3.04755	-2.85413	-2.84554	-2.891	-2.90937	-3.20755

Table B.3: Optimized trial function parameters for the SAPT2-retarded(97) potential in pure helium clusters.

	He_2	He_3	He_4	He_5	He_6	He_7	He_8	He_9	He_{10}	He_{20}
a_0	-0.969758	-0.649198	-0.259291	-0.206768	-0.130016	-0.129461	-0.11544	-0.109227	-0.109673	-0.115441
a_1	-0.00704896	-0.0308469	-0.0487016	-0.0434617	-0.0435104	-0.043574	-0.0411078	-0.0408033	-0.0404212	-0.0423227
b	1.0	1.0	0.993307	1.01054	1.00327	0.971732	0.964324	0.937381	0.917324	0.71875
α	0.846779	0.818231	0.937369	0.946901	0.965921	0.969517	0.970329	0.973689	0.969323	0.962697
β	4.07801e-06	-0.000139377	0.000315294	0.000300735	0.000247179	0.000250068	0.000238463	0.000262633	0.000264392	0.000287658
t_0	-245.099	-209.937	-239.522	-247.46	-250.633	-252.74	-251.576	-255.229	-255.02	-254.857
t_1	59.3337	50.1871	42.1065	43.5725	41.2331	41.2626	41.0733	41.3413	41.7439	42.7932
t_2	-5.22275	-4.21957	-2.8833	-3.07715	-2.77426	-2.86182	-2.8532	-2.89193	-2.90821	-3.19927

Table B.4: Optimized Mg-He trial function parameters for the CCSDT(Hinde) & SAPT2-retarded(97) potentials in singly doped $MgHe_n$ clusters.

	MgHe	$MgHe_2$	$MgHe_3$	$MgHe_4$	$MgHe_5$	$MgHe_6$	$MgHe_7$	$MgHe_8$	$MgHe_9$	$MgHe_{10}$	$MgHe_{20}$
a_0	0.0	0.0	0.0	-0.0159977	-0.0128084	-0.0122201	-0.0114488	-0.0107389	-0.0104444	-0.0101448	-0.00791426
a_1	-0.270121	-0.301119	-0.289189	-0.278073	-0.270623	-0.263391	-0.254985	-0.250296	-0.248481	-0.248919	-0.214723
b	1.0	0.944531	0.934546	0.92308	0.915755	0.911783	0.903789	0.897323	0.888944	0.880683	0.885098
α	0.71758	0.712703	0.711112	0.710733	0.710954	0.711602	0.712625	0.713465	0.713855	0.714578	0.71827
β	0.0	0.0	0.0	2.74269e-05	2.56969e-05	2.48694e-05	2.33039e-05	2.18729e-05	2.14326e-05	2.12128e-05	2.07824e-05
t_0	-552.962	-555.837	-553.395	-556.08	-558.459	-561.32	-564.243	-570.461	-575.938	-581.408	-611.833
t_1	3.32853	1.95335	1.34094	1.03827	0.971908	0.921617	0.8258	0.774366	0.75292	0.730992	0.680174
t_2	-2.27996	-1.67123	-1.57124	-1.41877	-1.41706	-1.41503	-1.38033	-1.33267	-1.33935	-1.34822	-1.32947

Table B.5: Optimized He-He trial function parameters for the CCSDT(Hinde) & SAPT2-retarded(97) potentials in singly doped $MgHe_n$ clusters.

	$MgHe_2$	$MgHe_3$	$MgHe_4$	$MgHe_5$	$MgHe_6$	$MgHe_7$	$MgHe_8$	$MgHe_9$	$MgHe_{10}$	$MgHe_{20}$
a_0	-0.139451	-0.128391	-0.125627	-0.120168	-0.114752	-0.106659	-0.0993346	-0.0951159	-0.0919446	-0.071389
a_1	-0.044532	-0.0450314	-0.0452103	-0.0430317	-0.0401898	-0.0396087	-0.038627	-0.0372635	-0.035823	-0.030555
b	1.00421	0.975138	0.944063	0.923861	0.915206	0.903139	0.887713	0.877565	0.866776	0.7231
α	0.96828	0.969797	0.973714	0.974835	0.976444	0.979138	0.984493	0.987865	0.990394	1.00545
β	0.00022561	0.000154194	0.000133979	0.000108419	0.00010975	0.00010323	9.47746e-05	8.29856e-05	9.20751e-05	8.80252e-05
t_0	-251.05	-252.36	-252.05	-252.372	-252.001	-252.445	-252.788	-253.334	-253.898	-263.732
t_1	40.2191	39.8432	39.1846	39.0704	38.8818	38.3987	37.7809	37.4943	37.3165	37.2014
t_2	-2.62959	-2.59528	-2.59217	-2.58771	-2.59622	-2.62428	-2.60368	-2.63178	-2.64263	-2.5628

Table B.6: Optimized Ca-He trial function parameters for the CCSDT(Hinde) & SAPT2-retarded(97) potentials in singly doped $CaHe_n$ clusters.

	CaHe	$CaHe_2$	$CaHe_3$	$CaHe_4$	$CaHe_5$	$CaHe_6$	$CaHe_7$	$CaHe_8$	$CaHe_9$	$CaHe_{10}$	$CaHe_{20}$
a_0	0.0419914	0.0372128	0.035858	0.03595	0.0369462	0.0371405	0.0386373	0.0382596	0.0376924	0.0372191	0.0331632
a_1	-0.225379	-0.212703	-0.205651	-0.195902	-0.188716	-0.18344	-0.178459	-0.17492	-0.171342	-0.169095	-0.171705
b	1.0	0.985207	0.969082	0.958795	0.945571	0.934081	0.924768	0.911808	0.901313	0.890152	0.788119
α	0.586605	0.587153	0.587	0.58772	0.588664	0.589946	0.591184	0.593328	0.59504	0.596724	0.609513
β	-1.55955e-4	-1.37286e-4	-1.2636e-4	-1.10297e-4	-1.05155e-4	-9.41496e-5	-1.00852e-4	-9.62592e-5	-9.06104e-5	-8.71407e-5	-6.91837e-5
t_0	-518.738	-520.781	-519.739	-520.514	-518.331	-522.253	-525.225	-530.645	-536.74	-541.5	-594.734
t_1	0.0	-0.00857286	-0.00802649	-0.00779543	-0.00738299	-0.00697277	-0.0068631	-0.00647945	-0.00612548	-0.00580502	-0.00437136
t_2	0.0	8.44996e-4	7.77637e-4	7.3281e-4	6.70455e-4	6.26342e-4	5.83699e-4	5.42635e-4	5.13566e-4	4.82185e-4	3.71324e-4

Table B.7: Optimized He-He trial function parameters for the CCSDT(Hinde) & SAPT2-retarded(97) potentials in singly doped $CaHe_n$ clusters.

	$CaHe_2$	$CaHe_3$	$CaHe_4$	$CaHe_5$	$CaHe_6$	$CaHe_7$	$CaHe_8$	$CaHe_9$	$CaHe_{10}$	$CaHe_{20}$
a_0	-0.164824	-0.160057	-0.153628	-0.142359	-0.132037	-0.121	-0.108131	-0.0976365	-0.084581	-0.0632033
a_1	-0.0484271	-0.0490244	-0.0479678	-0.0465892	-0.0452168	-0.0431566	-0.0416912	-0.0403975	-0.0393236	-0.0312216
b	0.993548	0.971763	0.95205	0.936158	0.920237	0.908409	0.898964	0.892782	0.886649	0.814617
α	0.974397	0.974282	0.976173	0.979764	0.984595	0.988026	0.991557	0.994002	0.997332	0.997789
β	0.000187121	0.0001772	0.000158725	0.000106955	0.000133173	0.000122587	0.000113541	0.000103707	9.53581e-05	9.68657e-05
t_0	-253.675	-255.02	-254.672	-255.614	-255.453	-255.491	-255.576	-255.35	-256.071	-253.821
t_1	39.7166	39.4222	38.9603	38.7418	38.1741	37.8787	37.3724	36.9597	36.6042	35.2286
t_2	-2.50296	-2.47384	-2.53439	-2.58281	-2.62724	-2.65031	-2.67272	-2.66057	-2.65089	-2.5341

Appendix C

He(CM)-Ca and He(CM)-Mg Density Distribution Functions

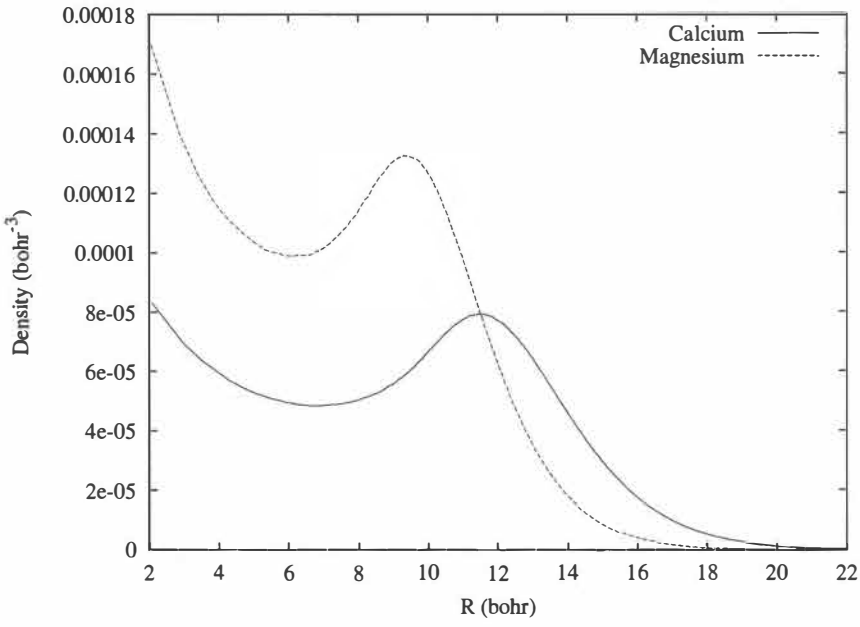


Figure C.1: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_2$ and $CaHe_2$.

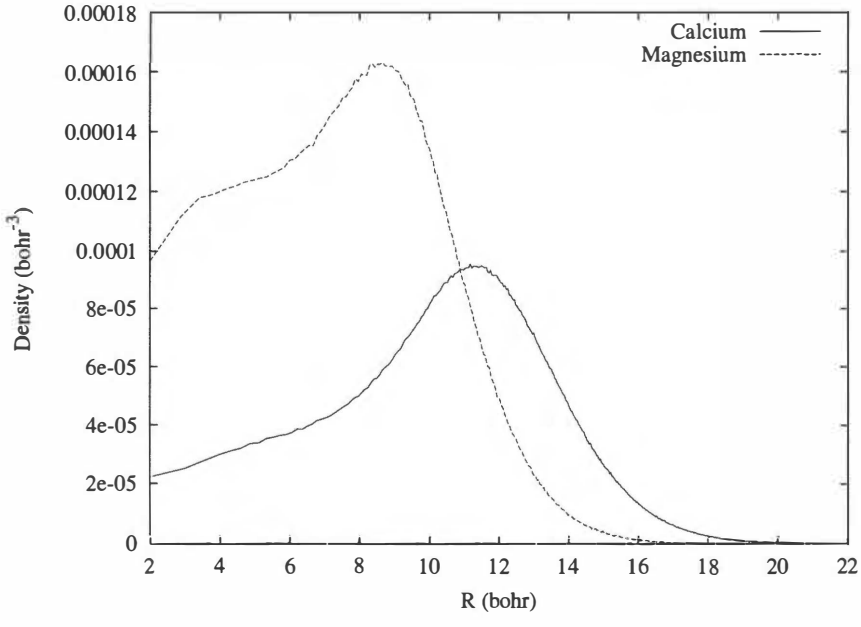


Figure C.2: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_3$ and $CaHe_3$.

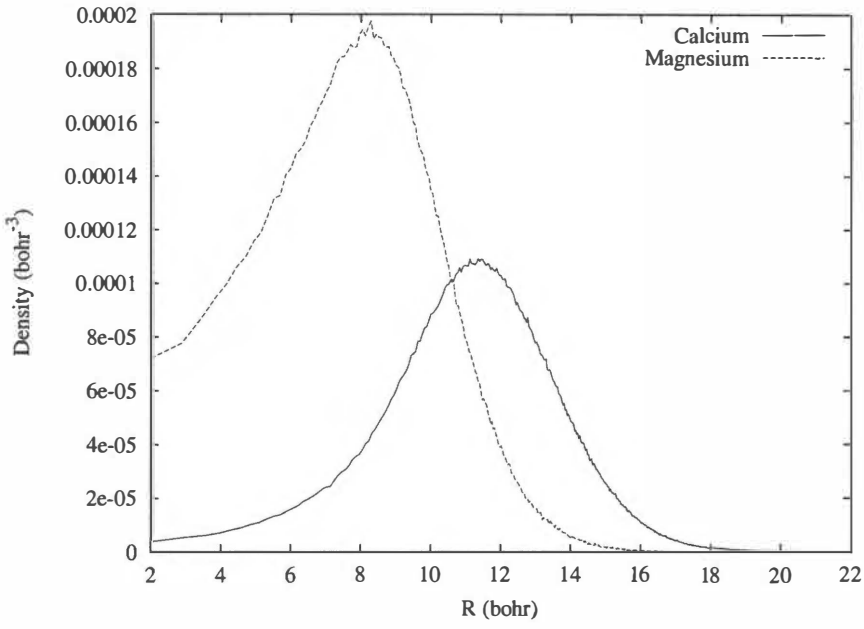


Figure C.3: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_4$ and $CaHe_4$.

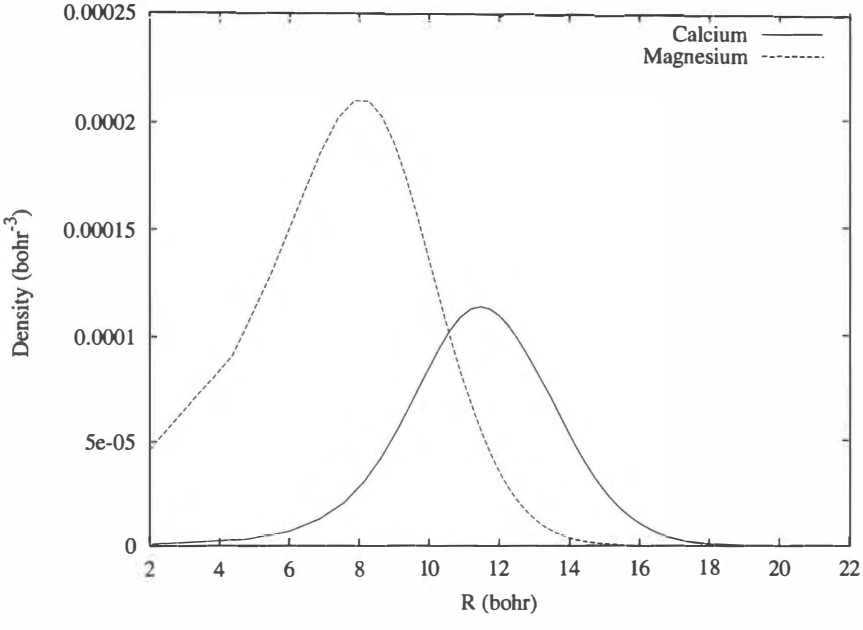


Figure C.4: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_5$ and $CaHe_5$.

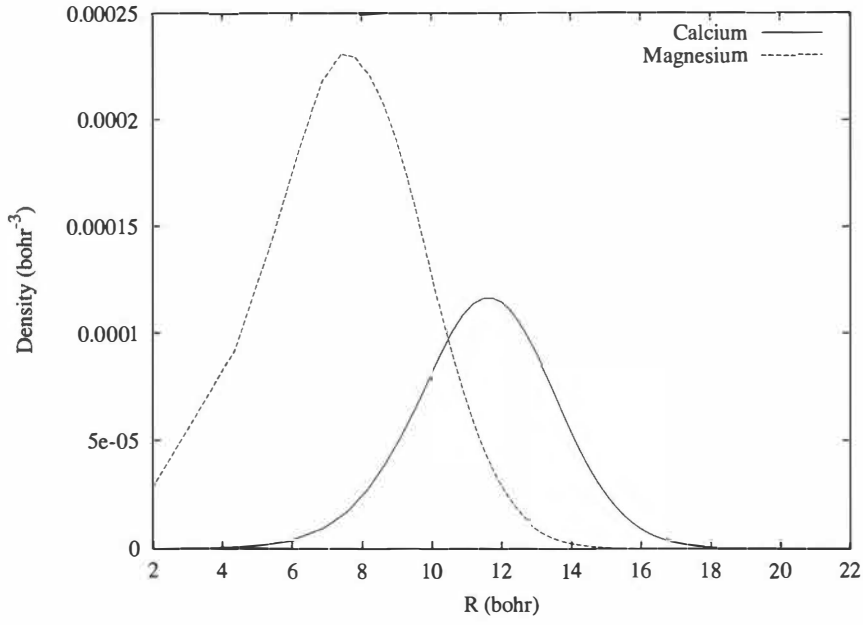


Figure C.5: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_6$ and $CaHe_6$.

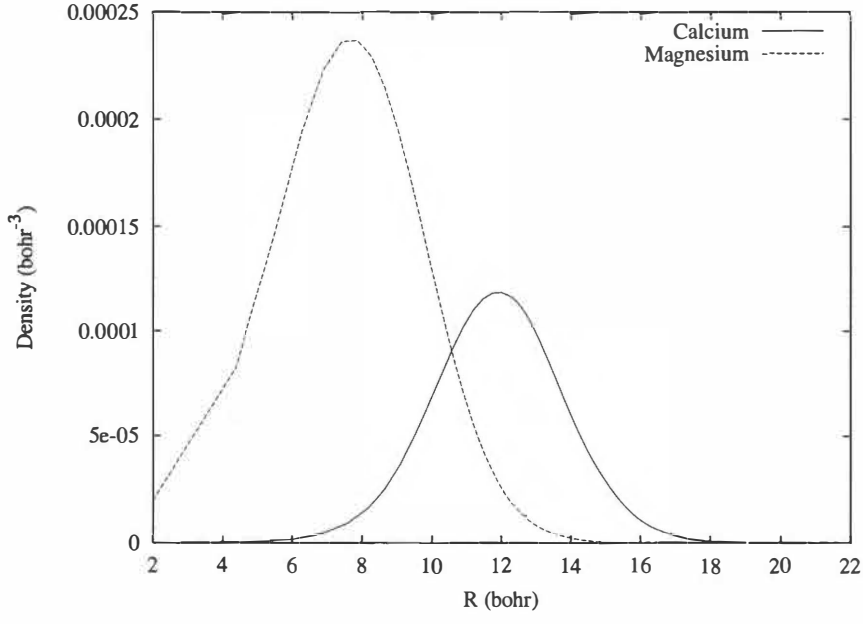


Figure C.6: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_7$ and $CaHe_7$.

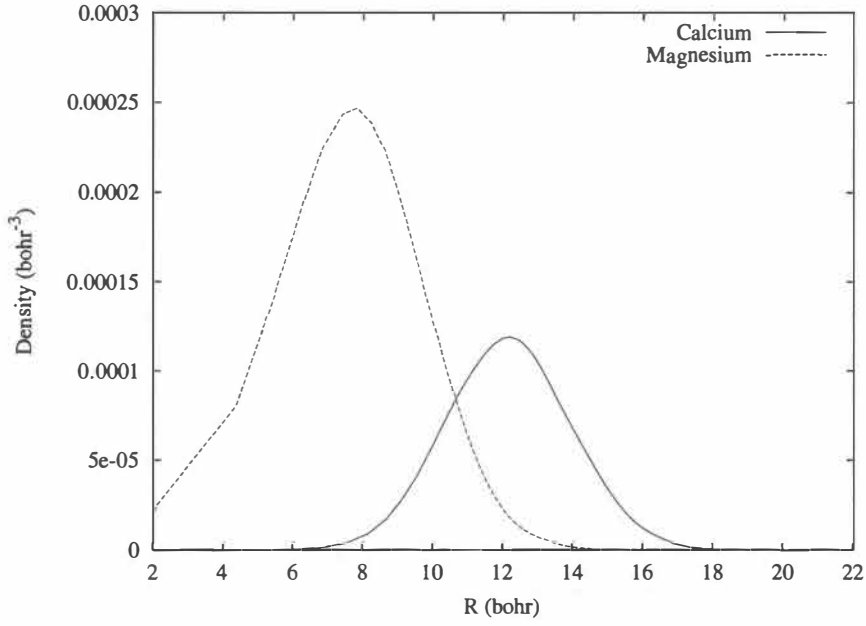


Figure C.7: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_8$ and $CaHe_8$.

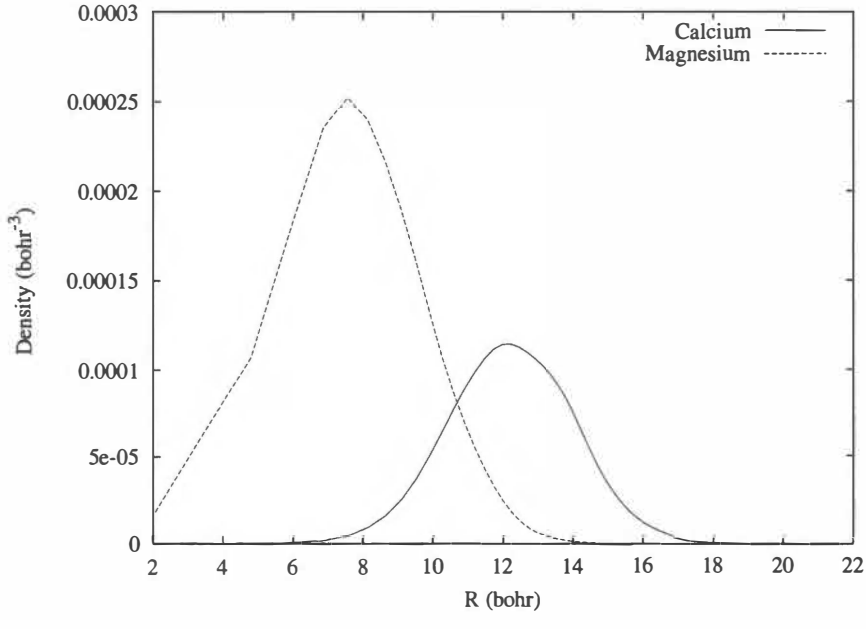


Figure C.8: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_9$ and $CaHe_9$.

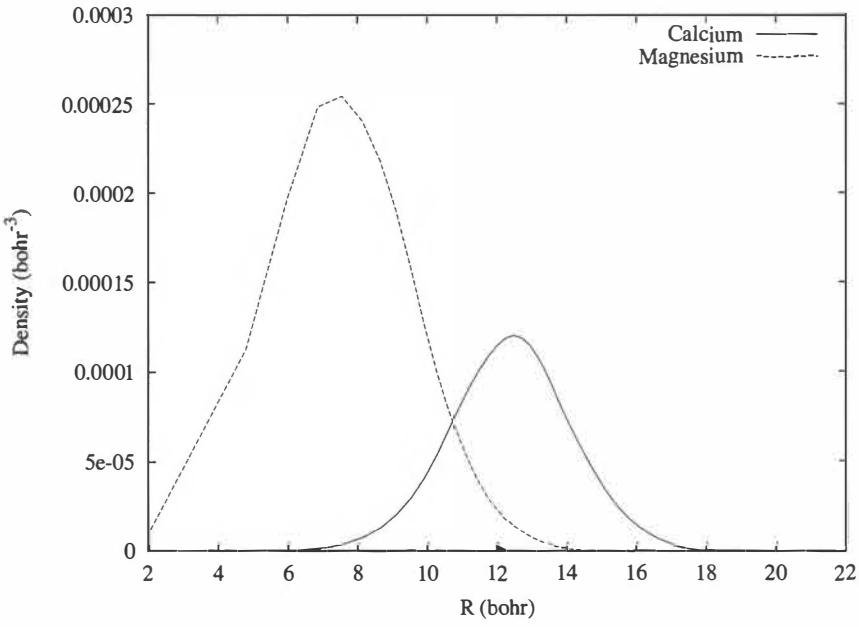


Figure C.9: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_{10}$ and $CaHe_{10}$.

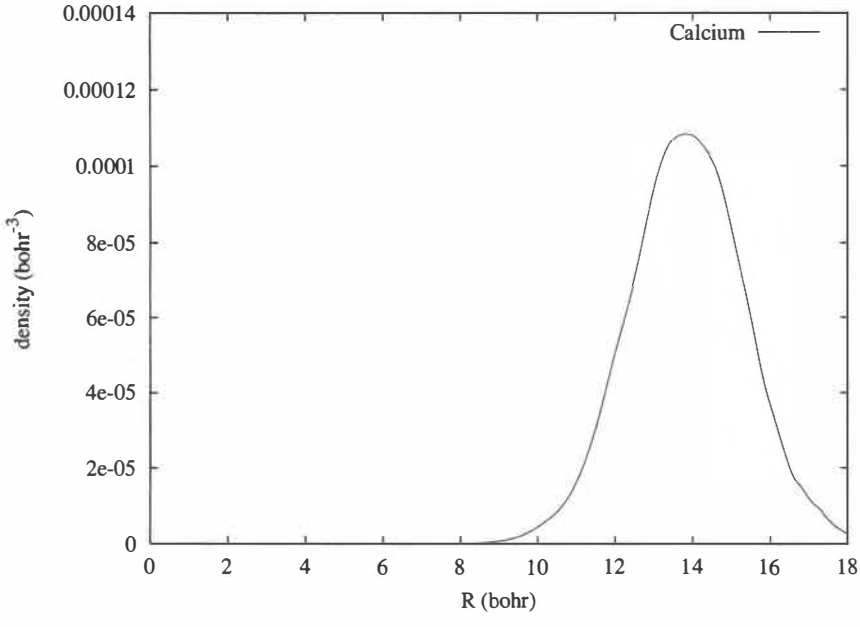


Figure C.10: Profile of Ca and Mg density from the mass center of the helium moiety in $MgHe_{20}$ and $CaHe_{20}$.

Appendix D

He-He-Ca and He-He-Mg Density Distribution Functions

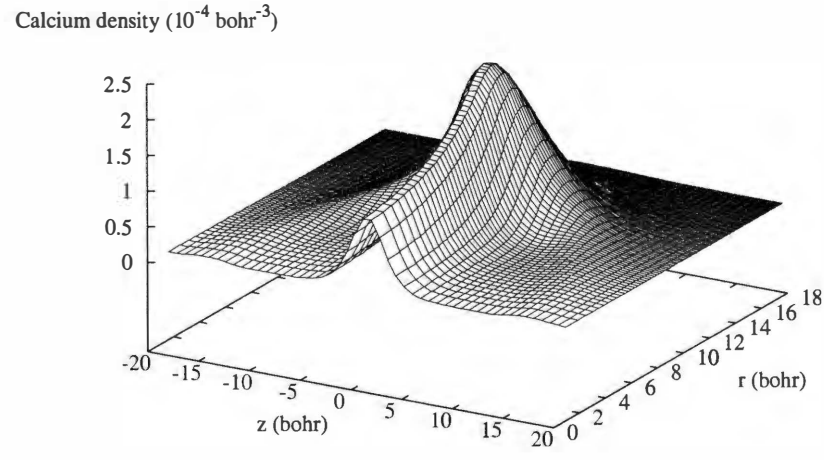


Figure D.1: Distribution of calcium density around helium pairs in $CaHe_2$.

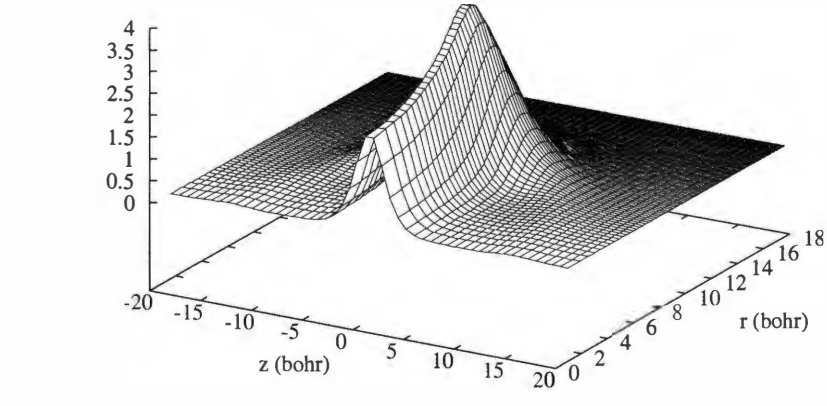


Figure D.2: Distribution of magnesium density around helium pairs in $MgHe_2$.

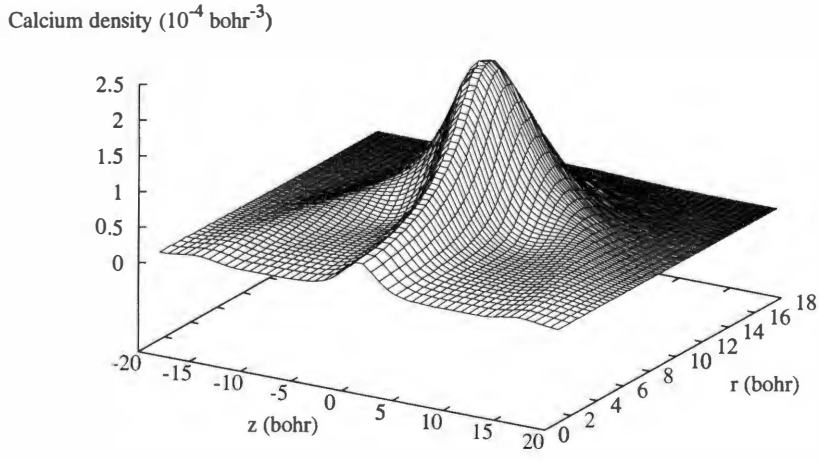


Figure D.3: Distribution of calcium density around helium pairs in $CaHe_3$.

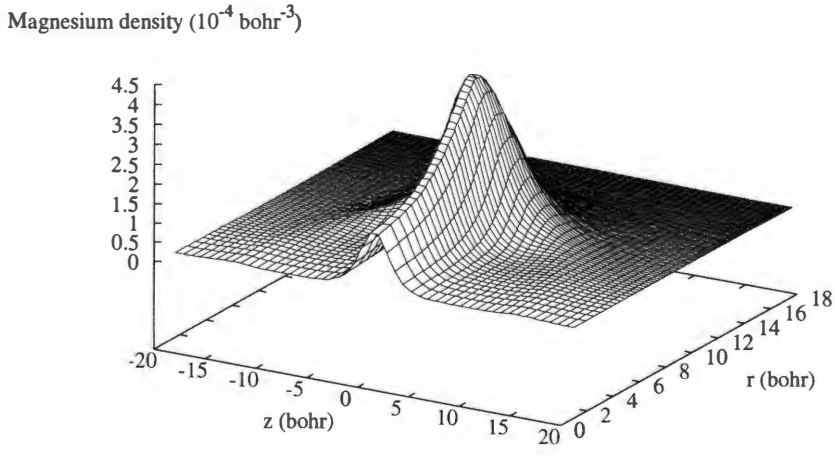


Figure D.4: Distribution of magnesium density around helium pairs in $MgHe_3$.

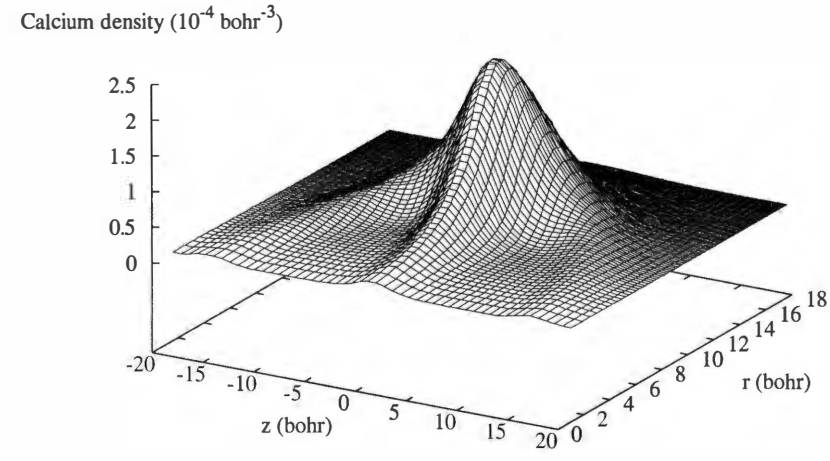


Figure D.5: Distribution of calcium density around helium pairs in $CaHe_4$.

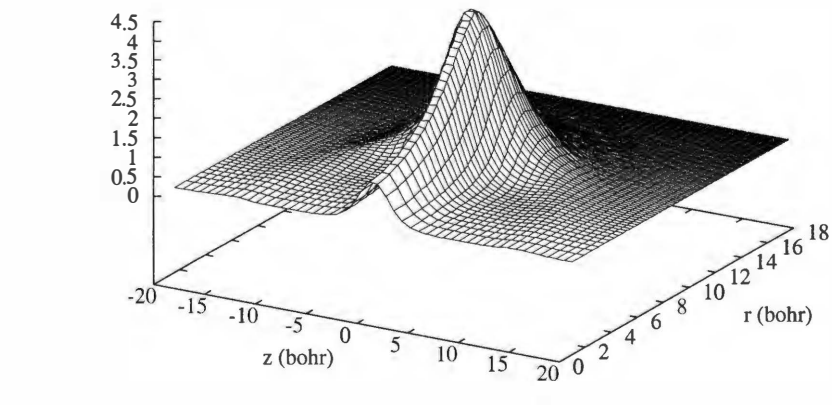


Figure D.6: Distribution of magnesium density around helium pairs in $MgHe_4$.

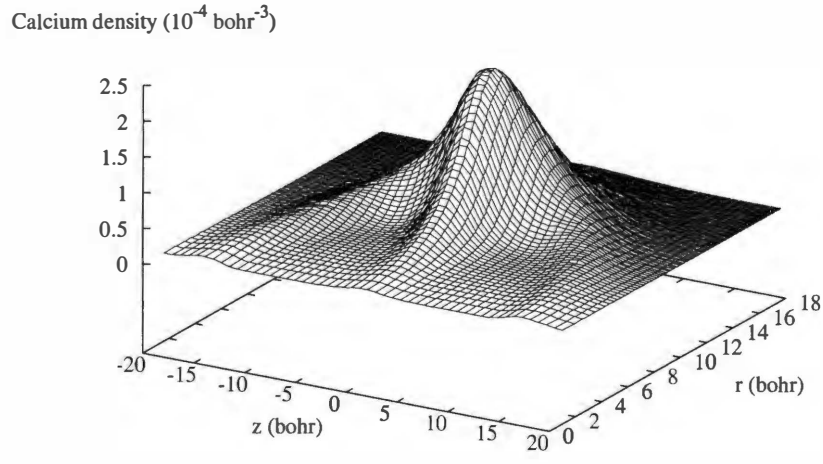


Figure D.7: Distribution of calcium density around helium pairs in $CaHe_5$.

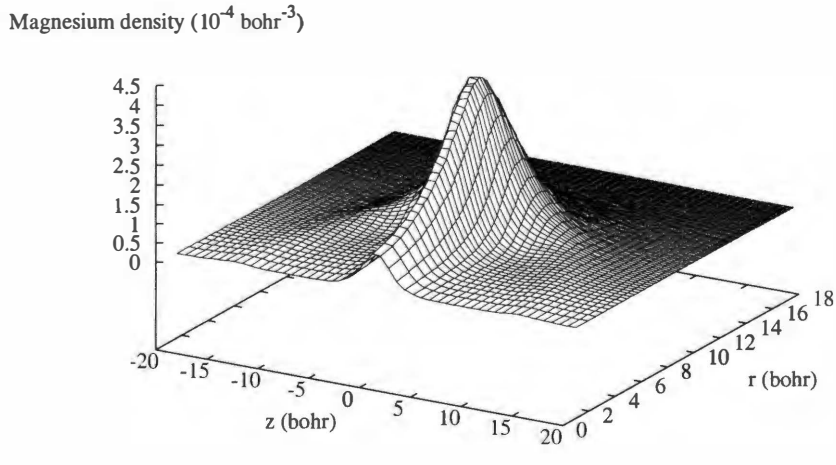


Figure D.8: Distribution of magnesium density around helium pairs in $MgHe_5$.

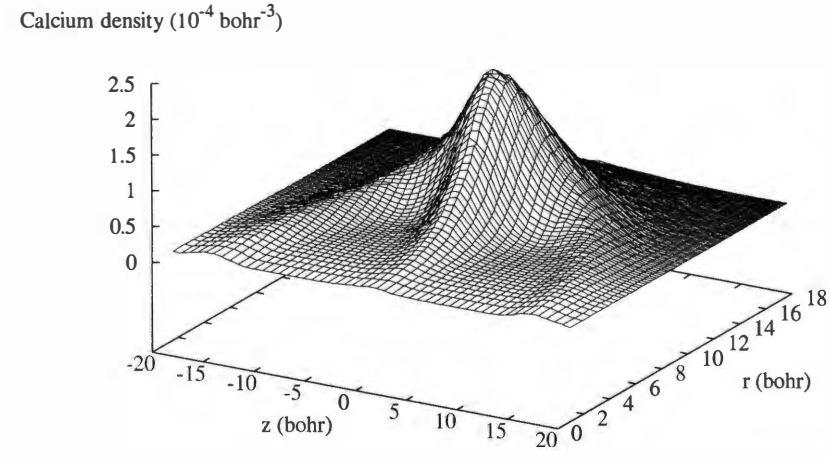


Figure D.9: Distribution of calcium density around helium pairs in $CaHe_6$.

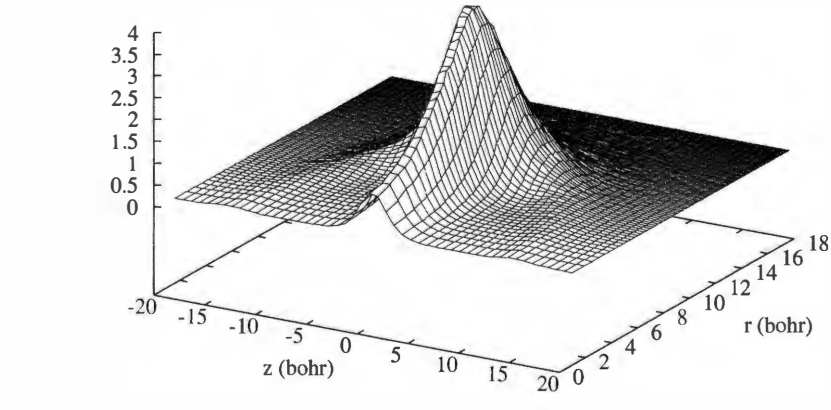


Figure D.10: Distribution of magnesium density around helium pairs in $MgHe_6$.

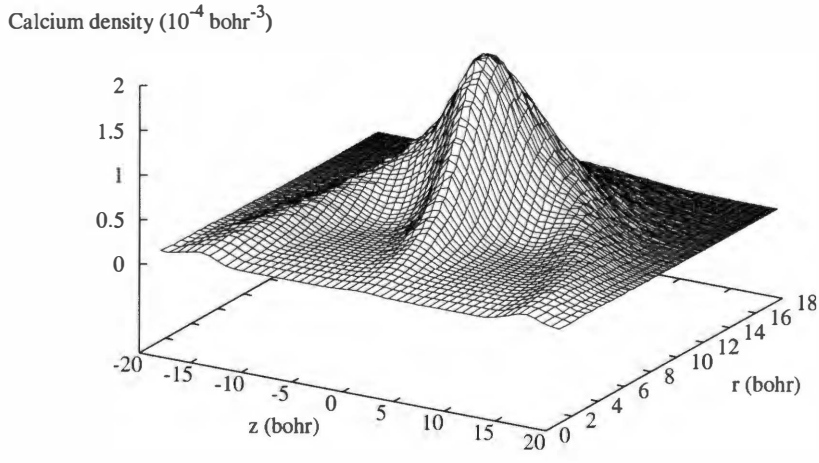


Figure D.11: Distribution of calcium density around helium pairs in $CaHe_7$.

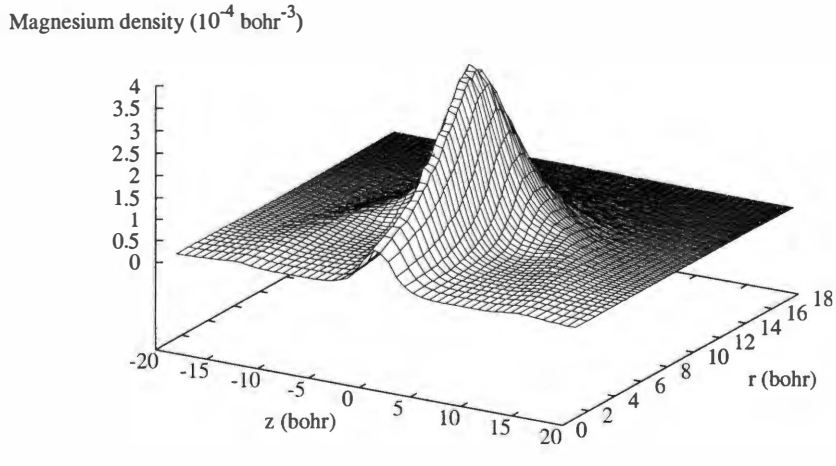


Figure D.12: Distribution of magnesium density around helium pairs in $MgHe_7$.

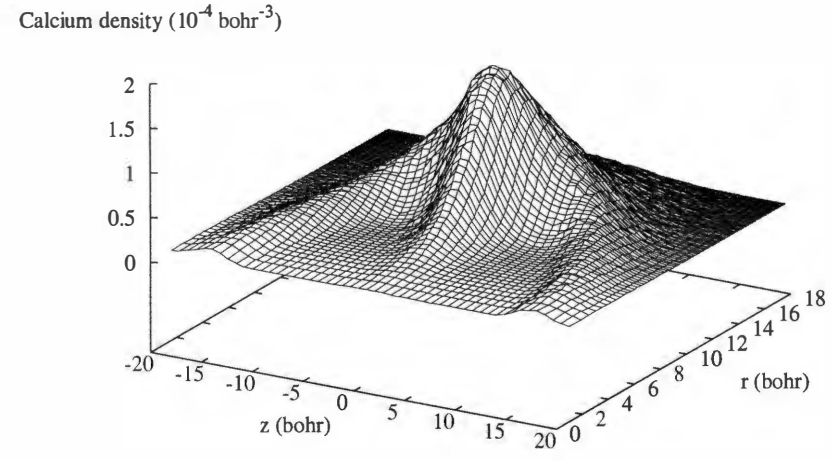


Figure D.13: Distribution of calcium density around helium pairs in $CaHe_8$.

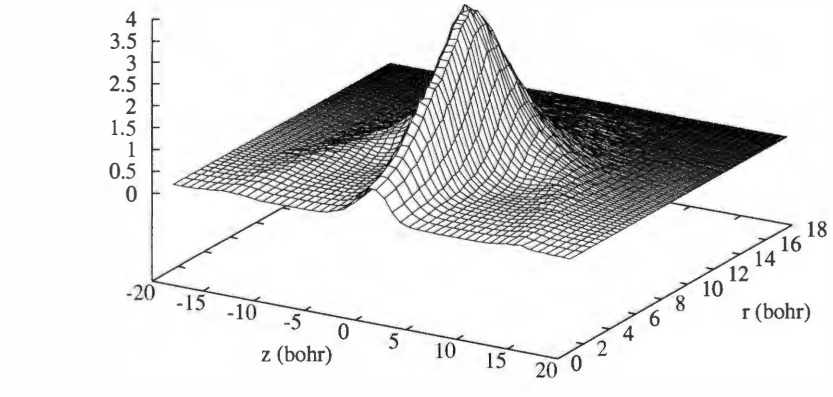


Figure D.14: Distribution of magnesium density around helium pairs in $MgHe_8$.

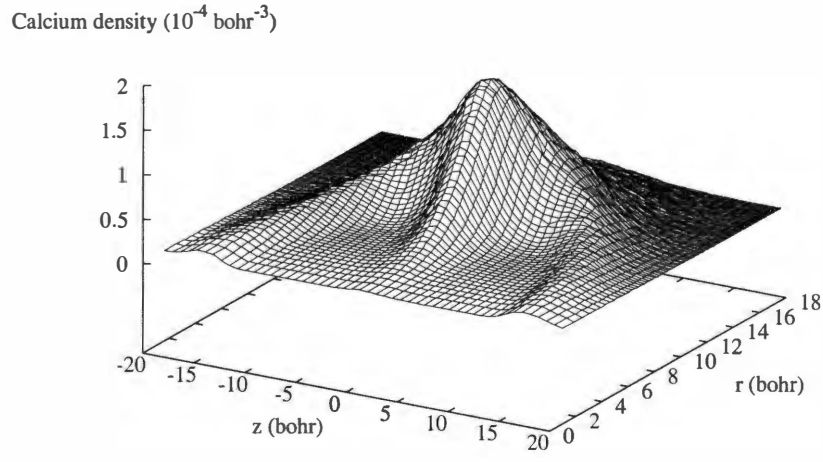


Figure D.15: Distribution of calcium density around helium pairs in $CaHe_9$.

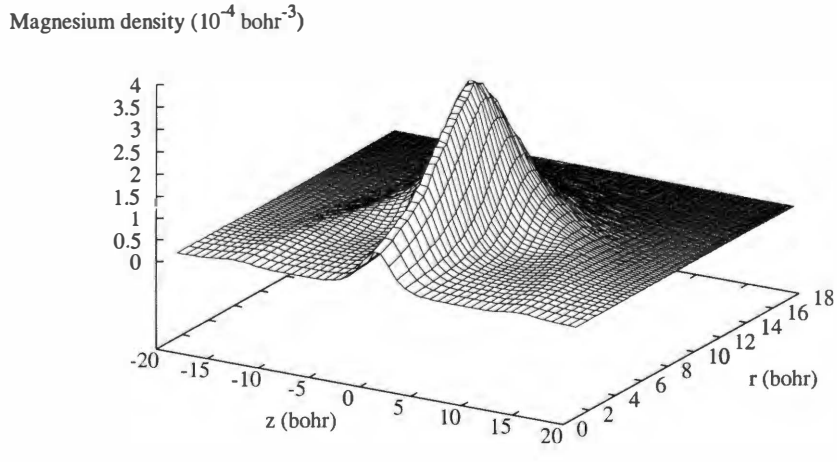


Figure D.16: Distribution of magnesium density around helium pairs in $MgHe_9$.

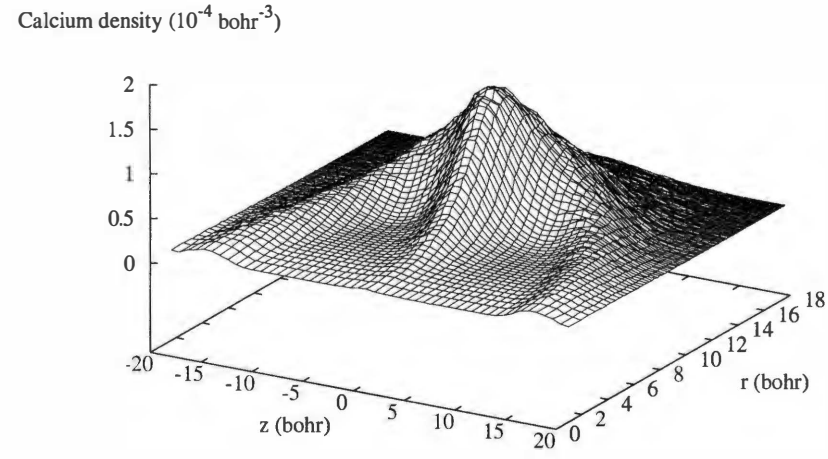


Figure D.17: Distribution of calcium density around helium pairs in $CaHe_{10}$.

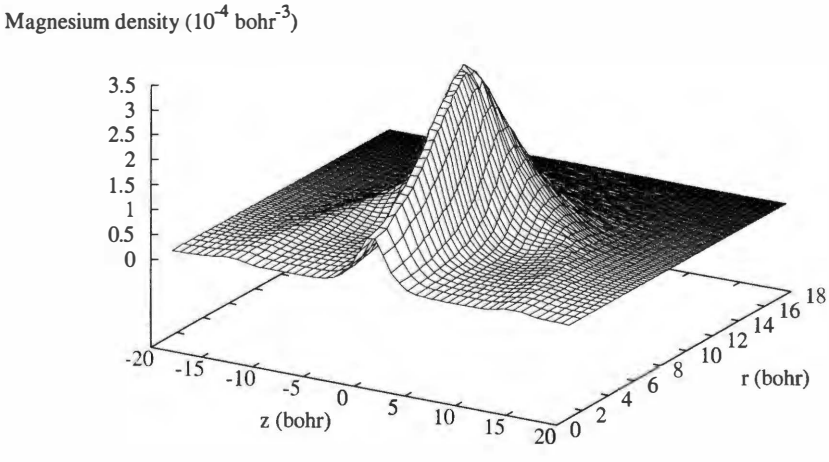


Figure D.18: Distribution of magnesium density around helium pairs in $MgHe_{10}$.

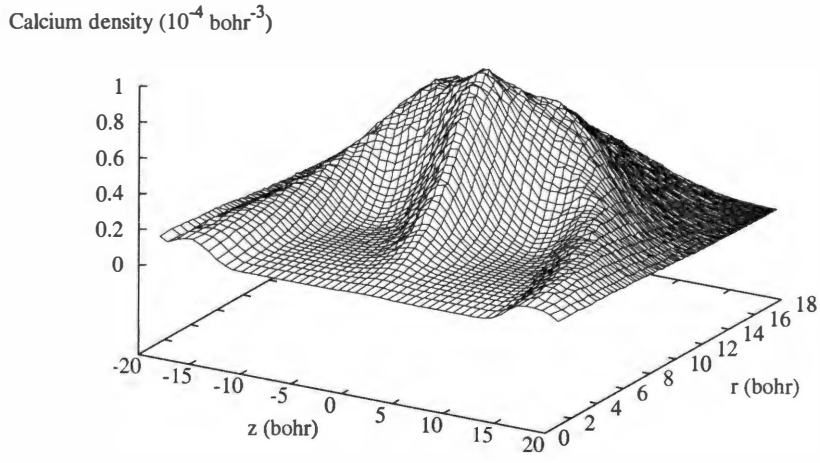


Figure D.19: Distribution of calcium density around helium pairs in $CaHe_{20}$.

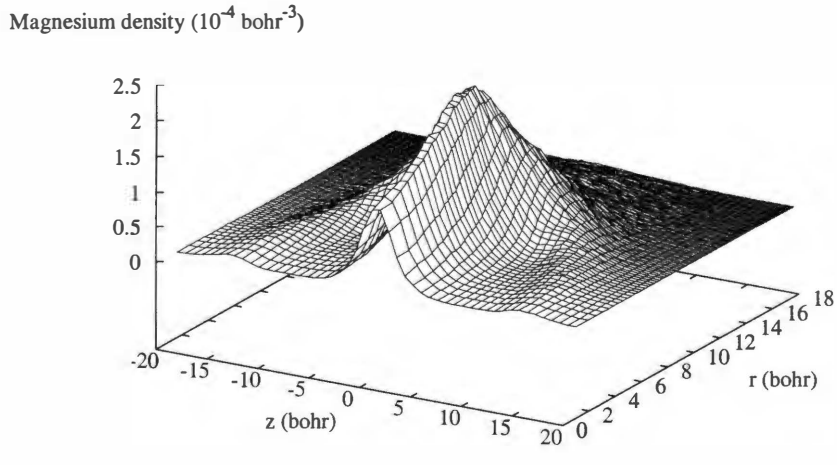


Figure D.20: Distribution of magnesium density around helium pairs in $MgHe_{20}$.

Appendix E

He(CM)-Ca-He and He(CM)-Mg-He Density Distribution Functions

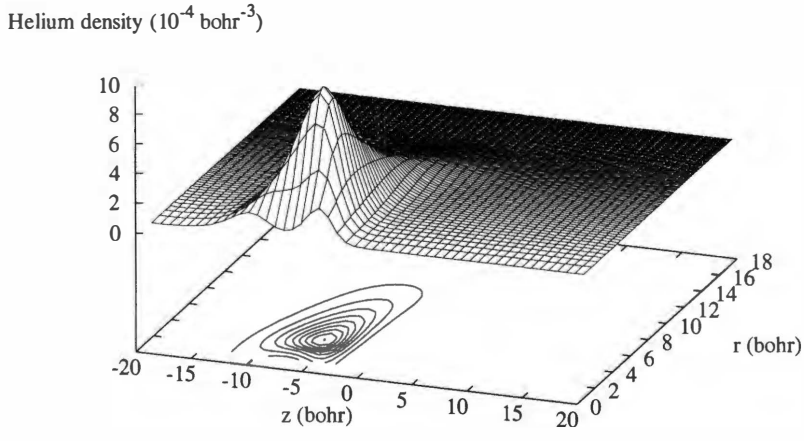


Figure E.1: Distribution of helium density in $CaHe_2$ around the internal axis connecting the calcium atom and the helium center of mass.

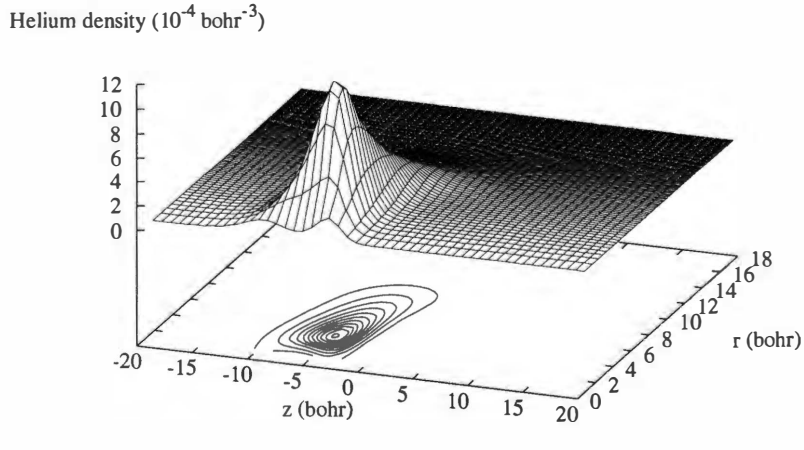


Figure E.2: Distribution of helium density in $MgHe_2$ around the internal axis connecting the magnesium atom and the helium center of mass.

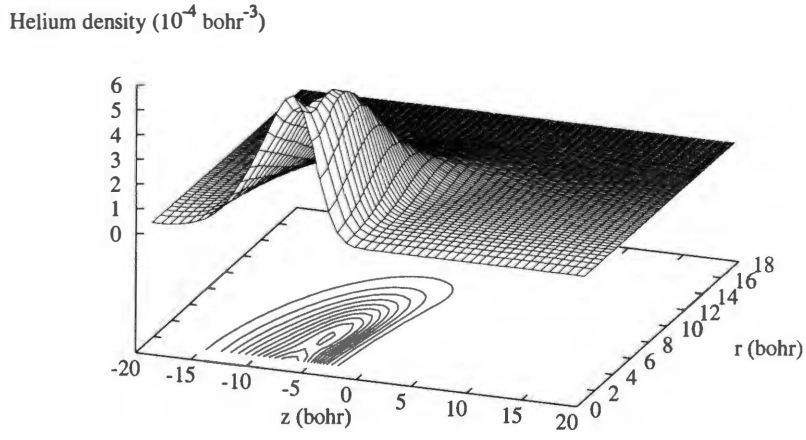


Figure E.3: Distribution of helium density in $CaHe_3$ around the internal axis connecting the calcium atom and the helium center of mass.

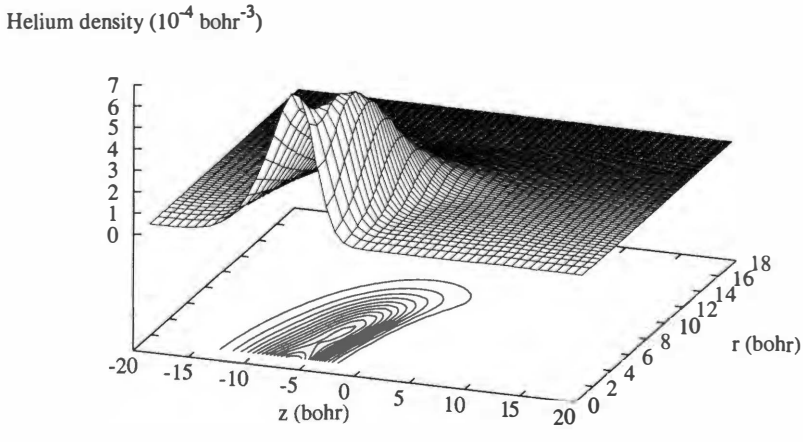


Figure E.4: Distribution of helium density in $MgHe_3$ around the internal axis connecting the magnesium atom and the helium center of mass.

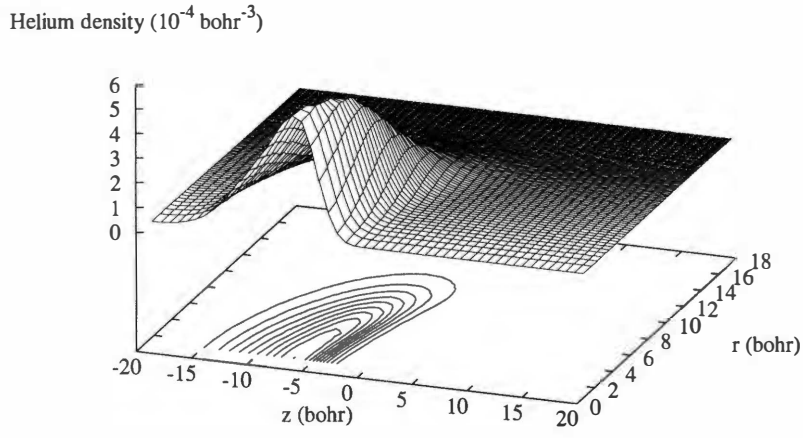


Figure E.5: Distribution of helium density in $CaHe_4$ around the internal axis connecting the calcium atom and the helium center of mass.

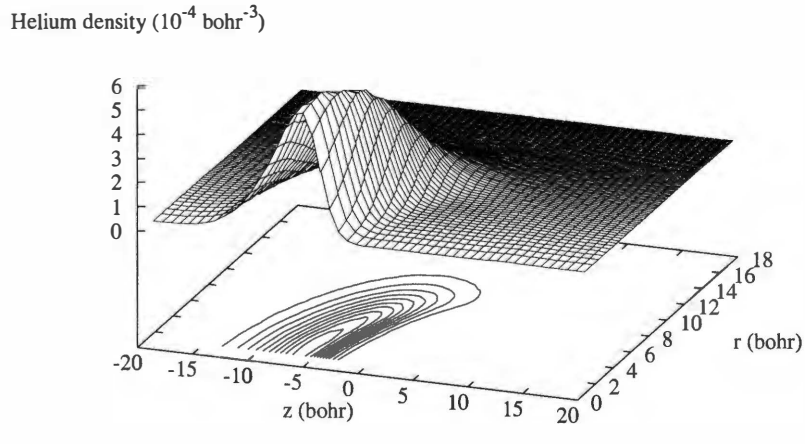


Figure E.6: Distribution of helium density in $MgHe_4$ around the internal axis connecting the magnesium atom and the helium center of mass.

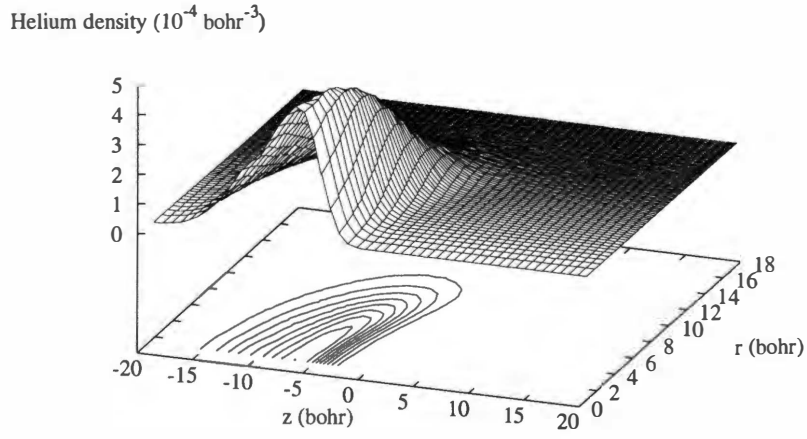


Figure E.7: Distribution of helium density in $CaHe_5$ around the internal axis connecting the calcium atom and the helium center of mass.

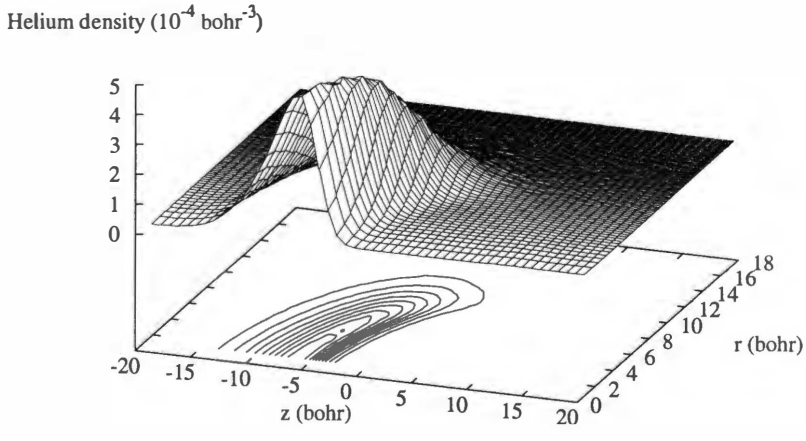


Figure E.8: Distribution of helium density in $MgHe_5$ around the internal axis connecting the magnesium atom and the helium center of mass.

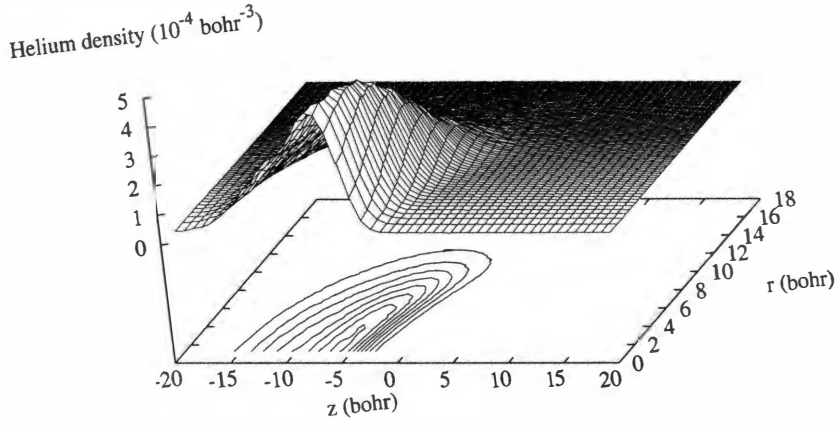


Figure E.9: Distribution of helium density in $CaHe_6$ around the internal axis connecting the calcium atom and the helium center of mass.

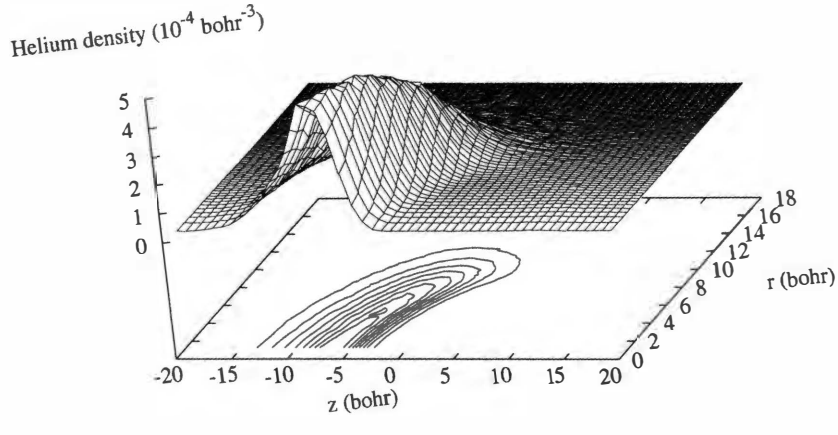


Figure E.10: Distribution of helium density in $MgHe_6$ around the internal axis connecting the magnesium atom and the helium center of mass.

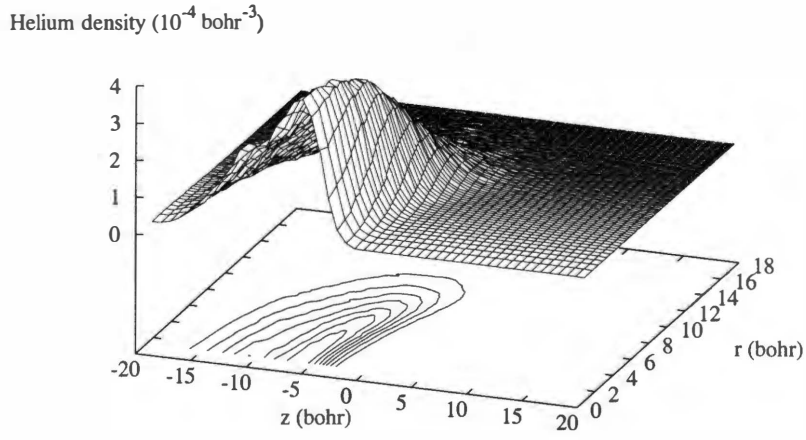


Figure E.11: Distribution of helium density in $CaHe_7$ around the internal axis connecting the calcium atom and the helium center of mass.

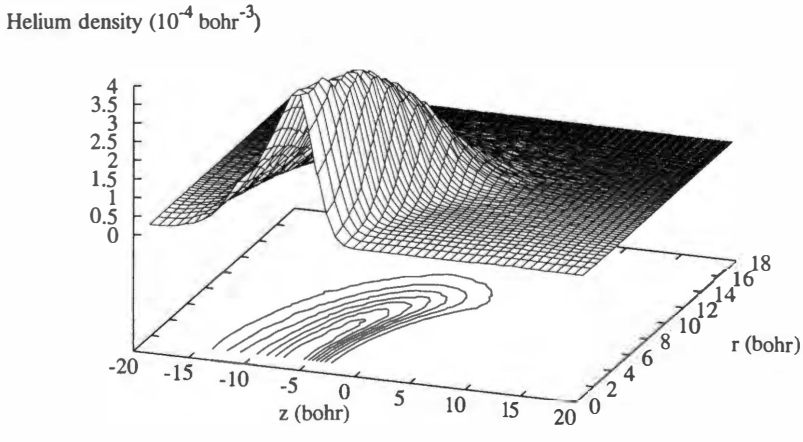


Figure E.12: Distribution of helium density in $MgHe_7$ around the internal axis connecting the magnesium atom and the helium center of mass.

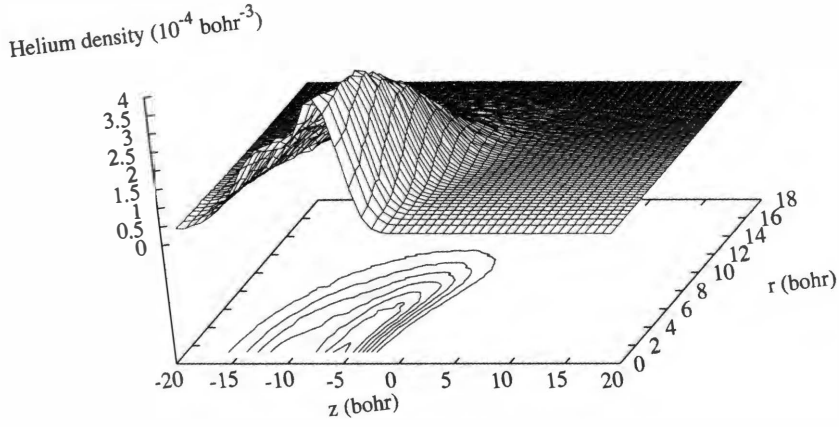


Figure E.13: Distribution of helium density in $CaHe_8$ around the internal axis connecting the calcium atom and the helium center of mass.

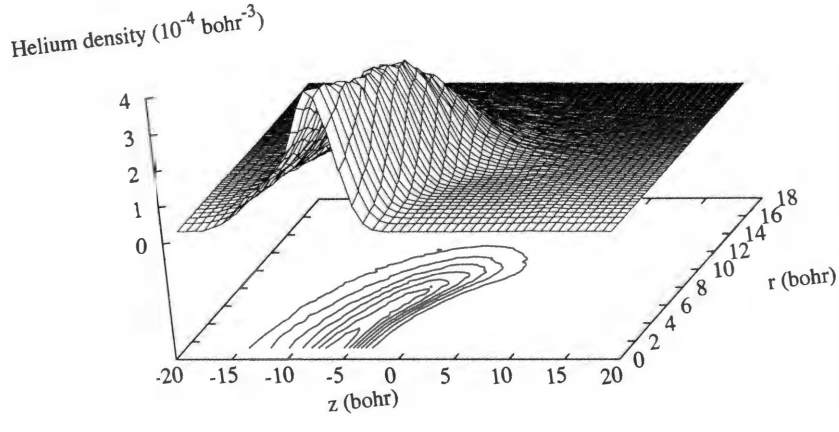


Figure E.14: Distribution of helium density in $MgHe_8$ around the internal axis connecting the magnesium atom and the helium center of mass.

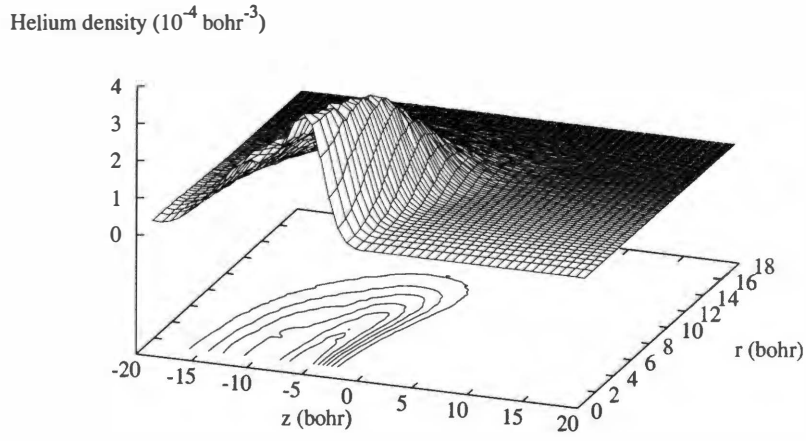


Figure E.15: Distribution of helium density in $CaHe_9$ around the internal axis connecting the calcium atom and the helium center of mass.

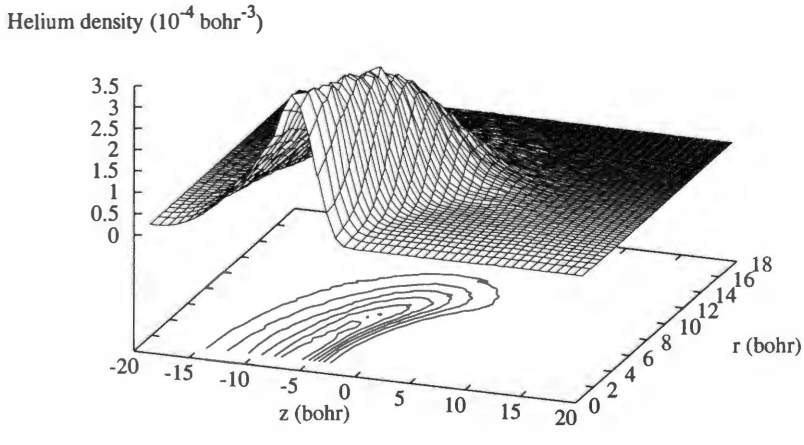


Figure E.16: Distribution of helium density in $MgHe_9$ around the internal axis connecting the magnesium atom and the helium center of mass.

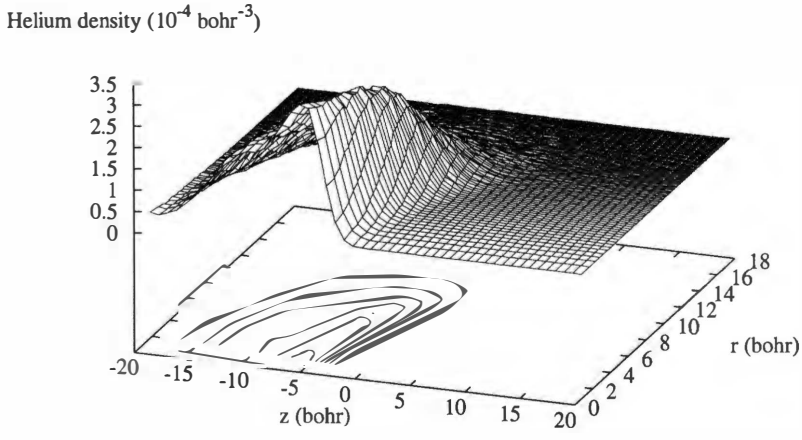


Figure E.17: Distribution of helium density in $CaHe_{10}$ around the internal axis connecting the calcium atom and the helium center of mass.

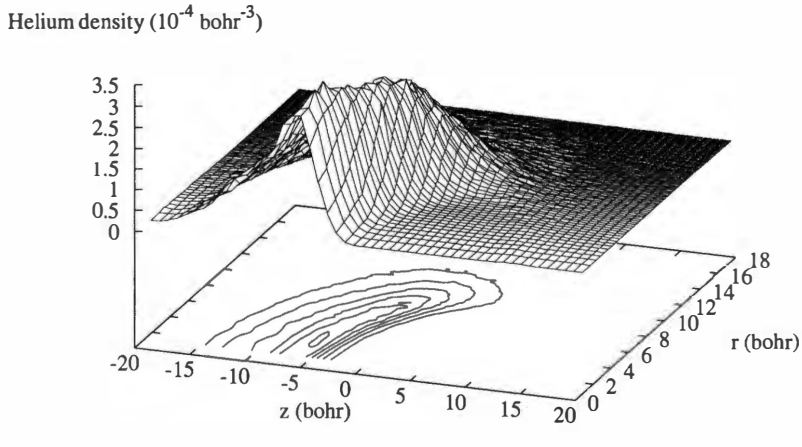


Figure E.18: Distribution of helium density in $MgHe_{10}$ around the internal axis connecting the magnesium atom and the helium center of mass.

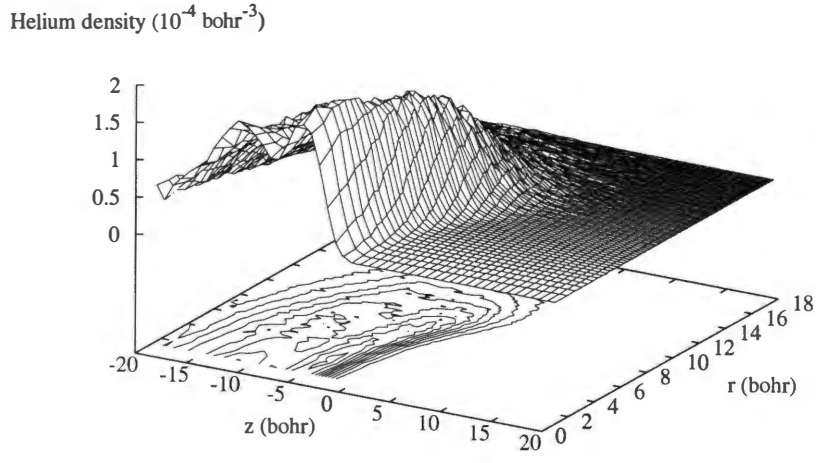


Figure E.19: Distribution of helium density in $CaHe_{20}$ around the internal axis connecting the calcium atom and the helium center of mass.

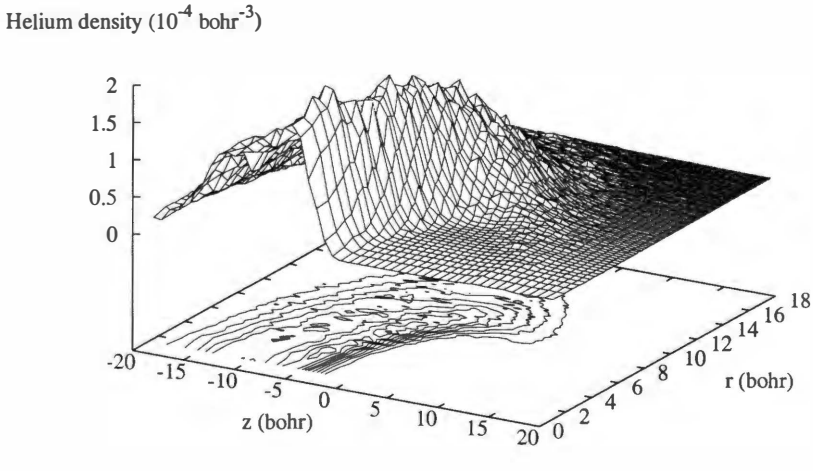


Figure E.20: Distribution of helium density in $MgHe_{20}$ around the internal axis connecting the magnesium atom and the helium center of mass.

VITA

Gary Lee Warren, Jr. was born in Knoxville, Tennessee on December 31, 1977. He graduated from Sumter High School in 1996 in the top 2% of his class. As an undergraduate, he attended Newberry College in Newberry, South Carolina. He graduated *summa cum laude* in 2000 with a Bachelor of Science in Chemistry and Mathematics and a minor in computer science. While at Newberry College, he met his wife, Lena and was married in May of 2000.

In the summer of 2000, Lee began working toward his Ph.D. at the University of Tennessee, working under Dr. Robert Hinde. While at the University of Tennessee, he was the recipient of the Eugene John Barber Fellowship in Chemistry and was awarded a National Science Foundation Graduate Research Fellowship. He graduated with a Ph.D. in May 2005 and plans to pursue a career in teaching and research.